# Identifying a neuroanatomical signature of schizophrenia, reproducible across sites and stages, using machine learning with structured sparsity

A. de Pierrefeu, T. Löfstedt, C. Laidi, F. Hadj-Selem, J. Bourgin, T. Hajek, F. Spaniel, M. Kolenic, P. Ciuciu, N. Hamdani, et al.

HAL Id: cea-01883283

https://hal-cea.archives-ouvertes.fr/cea-01883283

Submitted on 27 Sep 2018

# Identifying a neuroanatomical signature of schizophrenia, reproducible across sites and stages, using machine learning with structured sparsity

de Pierrefeu A, Löfstedt T, Laidi C, Hadj-Selem F, Bourgin J, Hajek T, Spaniel F, Kolenic M, Ciuciu P, Hamdani N, Leboyer M, Fovet T, Jardri R, Houenou J, Duchesnay E. Identifying a neuroanatomical signature of schizophrenia, reproducible across sites and stages, using machine learning with structured sparsity

**Objective:** Structural MRI (sMRI) increasingly offers insight into abnormalities inherent to schizophrenia. Previous machine learning applications suggest that individual classification is feasible and reliable and, however, is focused on the predictive performance of the clinical status in cross-sectional designs, which has limited biological perspectives. Moreover, most studies depend on relatively small cohorts or single recruiting site. Finally, no study controlled for disease stage or medication's effect. These elements cast doubt on previous findings' reproducibility.

**Method:** We propose a machine learning algorithm that provides an interpretable brain signature. Using large datasets collected from 4 sites (276 schizophrenia patients, 330 controls), we assessed cross-site prediction reproducibility and associated predictive signature. For the first time, we evaluated the predictive signature regarding medication and illness duration using an independent dataset of first-episode patients.

**Results:** Machine learning classifiers based on neuroanatomical features yield significant intersite prediction accuracies (72%) together with an excellent predictive signature stability. This signature provides a neural score significantly correlated with symptom severity and the extent of cognitive impairments. Moreover, this signature demonstrates its efficiency on first-episode psychosis patients (73% accuracy).

**Conclusion:** These results highlight the existence of a common neuroanatomical signature for schizophrenia, shared by a majority of patients even from an early stage of the disorder.

**A. de Pierrefeu**[1] (iD)**, T. Löfstedt**[2]**,
C. Laidi**[1,3,4,5]**, F. Hadj-Selem**[6]**,
J. Bourgin**[7,8]**, T. Hajek**[9,10]**,
F. Spaniel**[10] (iD)**, M. Kolenic**[10]
**P. Ciuciu**[1,11] (iD)**, N. Hamdani**[3,4,5]**,
M. Leboyer**[3,4,5]**, T. Fovet**[12,13] (iD)**,
R. Jardri**[11,12,13] (iD)**, J.
Houenou**[1,3,4,5]**, E. Duchesnay**[1] (iD)

[1]NeuroSpin, CEA, Gif-sur-Yvette, France, [2]Department of Radiation Sciences, Umeå University, Umeå, Sweden, [3]Institut National de la Santé et de la Recherche Médicale (INSERM), U955, Institut Mondor de Recherche Biomédicale, Psychiatrie Translationnelle, [4]Fondation Fondamental, [5]Pôle de Psychiatrie, Assistance Publique–Hôpitaux de Paris (AP-HP), Faculté de Médecine de Créteil, DHU PePsy, Hôpitaux Universitaires Mondor, Créteil, [6]Energy Transition Institute: VeDeCoM, Versailles, [7]Department of Psychiatry, Louis-Mourier Hospital, AP-HP, Colombes, [8]INSERM U894, Centre for Psychiatry and Neurosciences, Paris, France, [9]Department of Psychiatry, Dalhousie University, Halifax, NS, Canada, [10]National Institute of Mental Health, Klecany, Czech Republic, [11]INRIA, CEA, Parietal team, University of Paris-Saclay, [12]Laboratoire de Sciences Cognitives et Sciences Affectives (SCALab-PsyCHIC), CNRS UMR 9193, University of Lille, and [13]Pôle de Psychiatrie, Unité CURE, CHU Lille, Lille, France

**Significant outcomes**
- Significant intersite prediction accuracy of clinical diagnosis based on sMRI.
- Identification of a robust and interpretable structural brain signature of schizophrenia.
- The predictive signature generalizes to the detection of patients at the early stage of the disorder.

**Limitations**
- At this stage, this does not imply that such predictive models are able to distinguish patients with various psychiatric conditions.
- A minority of patients do not present such brain abnormalities, which directly questions the need for a disorder stratification.

## Introduction

Schizophrenia is a disabling chronic mental disorder characterized by various symptoms such as hallucinations, delusions as well as impairments in high-order cognitive functions. By taking advantage of non-invasive imaging approaches, psychoradiology (https://radiopaedia.org/articles/psychoradiology), a subfield of radiology, is increasingly of clinical importance in guiding diagnostic and therapeutic decision-making in patients with mental disorders. Indeed, the development of magnetic resonance imaging (MRI) provides an effective and non-invasive approach to investigate the neuroanatomy of the brain. Specifically, structural MRI (sMRI) allows the study of structural changes in the brain and their relationship with the clinical diagnosis. Over the years, sMRI has been increasingly used to gain insights on the structural abnormalities inherent to the disorder and to identify brain regions where schizophrenia patients differ significantly from healthy controls (1). Unfortunately, there is indeed a wide overlap between brain imaging measurements in schizophrenia patients and the normal range (2). Thus, group analyses cannot be easily used to assist in the diagnosis process.

Recent progress in machine learning together with the availability of large datasets now pave the way for automatic detection of brain disorders, solely based on MRI data (3, 4). In the past, an extensive number of studies have focused on the prediction of schizophrenia based on neuroanatomical features (5–7). These studies uncovered relevant structural brain patterns that are different from controls and patients and that achieve a prediction at the individual level. Based on these structural discrepancies alone, classifiers reached various prediction performances ranging from 65% to 90% of accuracy.

However, to date, despite initial promising results, these studies have barely impacted clinical practice. Significant challenges still need to be tackled for translational implementation of such findings in psychiatry.

Schizophrenia is a complex and very heterogeneous disorder. Small size cohorts, typically composed of highly selected patients, suffer from a bias in the recruitment. They do not represent the full and broad cross-sectional spectrum of the disorder phenotype. Given this variability, a significant heterogeneity can be found in the effect sizes and patterns of brain differences across studies (8–10). To date, most studies recruited subjects scanned at a single acquisition site (i.e. the subjects were scanned at the same site, using similar scanner hardware and MRI protocols).

Such results are difficult to generalize to large-scale clinical settings, that is, with patients scanned at widely different locations (11). Validation on independent datasets is a more realistic approach to quantifying generalization accuracy. Consequently, multi-site populations are instrumental to achieve consistency and reproducibility in the results. To our knowledge, only a few studies have relied on a completely independent validation cohort to estimate the prediction performances of a classifier (12–14).

Leveraging those studies, we intend to further develop our findings along two different aspects. First, in the context of predictive signature discovery, it is crucial to understand the brain's structural patterns that underpin a prediction. Unfortunately, in most cases, despite accurate prediction performance achieved, classifiers still behave as a 'black box' model, not providing objective neuroanatomical markers thus ruling out the prospect of clinical application. We will therefore focus on the interpretability of such predictive patterns.

Second, we strive to filter-out chronic pharmaceutical treatments' impact on the brain. Given that the literature has consistently reported that some regions of the brain are affected by antipsychotic medication (15), our intention is to evaluate the generalization of the developed predictive models on subjects that are still in an early stage of the disease. Hence, we need to address the nonnegligible probability that previous classifiers rely heavily on the medication impacts over the brain rather than as 'true' markers of the disorder able to distinguish healthy individuals from those affected by schizophrenia.

### Aims of the study

Here, we validated automatic methods to classify schizophrenia using exclusively sMRI scans. We tested different sMRI-based features to assess intersite performance replicability using data from 606 subjects scanned at four distinct sites with no prior coordination. In addition, we investigated the interpretability of the obtained neuroanatomical predictive signature and its independence regarding medication. Finally, we tested the ability of our classifiers to generalize to an independent set of patients with first-episode psychosis.

## Methods

### Participants

Brain imaging data from four independent studies with no prior coordination were gathered in the

current analysis (http://schizconnect.org). The full dataset included 276 patients with strict schizophrenia, according to DSM-IV criteria, and 330 healthy controls. One additional independent set of healthy controls and patients with first-episode psychosis (FEP) was used for additional validation of the prediction performance.

Subjects provided informed consent to participate in their respective studies. Demographic details of all four datasets are summarized in Table 1. Information about the MRI acquisition protocols is gathered in Table S1. Prior to the analysis, raw MRI scans were visually controlled for motion and artefacts. A total of 57 scans did not survived this strict quality control and were excluded from further analysis. (Those subjects are not included in the 606 individuals detailed in Table 1.)

MRI preprocessing and features extraction

Prior to training classifiers, the first step was to compute samples from the structural MRI scans. The intention is to extract valuable information from the T1 images by converting the MRI scans into vectors of features. The choice of such features is crucial since it reflects different aspects of the brain anatomy. We retain three different types of features that are potentially powerful candidates to assist in the diagnosis of schizophrenia (16).

Grey matter voxel-based morphometry (VBM) maps were computed for each subject using SPM12 (17): segmentation, normalization and modulation steps are required. The sMRI images are first segmented into GM, WM and CSF. Then, in order to achieve spatial correspondence of voxels across subjects, all brain images are normalized into a common standard space using DARTEL nonlinear transformation (18). All the normalized images are finally modulated by the Jacobian of their transformation. This enables to preserve the quantity of tissue. To reduce computational cost, those images were subsampled to a $3 \times 3 \times 3$ mm resolution and the analysis was restricted to the voxels within an implicit brain mask, where the minimum value across subjects was larger than 1% and a standard deviation larger than $10^{-6}$. This produced 125 959 features representing the local grey matter volume at each voxel.

Vertex-based cortical thickness features were obtained with the FreeSurfer image analysis suite version 6.1 (http://surfer.nmr.mgh.harvard.edu/). Briefly, the preprocessing included removal of non-brain tissue, automated Talairach transformation, tessellation of the grey and white matter boundaries and surface deformation. A number of deformation procedures were performed in the data analysis pipeline, including surface inflation and registration to a spherical atlas. Intensity and continuity information from the entire three-dimensional images in the segmentation and deformation procedures were used to produce vertex-wise representations of cortical thickness (CT) in each vertex across the cortical mantle. This produced 299 862 features representing the cortical thickness at each vertex.

Last, 66 structural measurements of regions of interest (ROIs) were extracted with FreeSurfer which automatically parcellate the brain parenchyma in cortical and subcortical ROIs and compute

Table 1. Demographic and clinical characteristics of the dataset

| Datasets | Diagnosis | Patients | | | | Healthy controls | | |
|---|---|---|---|---|---|---|---|---|
| | | $n$ | Age (mean ± SD) | Gender (% F) | Clinical symptom scores type (mean ± SD) | $n$ | Age (mean ± SD) | Gender (% F) |
| Site 1 NUSDAST database | Schizophrenia strict (DSM-IV) | 118 | 33.95 ± 12.87 | 32 | SAPS (17.84 ± 15.2) SANS (21.15 ± 13.6) | 152 | 27.96 ± 12.58 | 54 |
| Site 2 COBRE database | Schizophrenia strict (DSM-IV) | 77 | 37.28 ± 13.56 | 16 | PANSS POS (14.92 ± 5.23) PANSS NEG (15.07 ± 5.21) | 87 | 38.33 ± 11.80 | 27 |
| Site 3 NMorphCH database | Schizophrenia strict (DSM-IV) | 42 | 32.78 ± 6.75 | 33 | SAPS (36.35 ± 28.74) SANS (41.97 ± 19.94) | 38 | 31.23 ± 7.94 | 52 |
| Site 4 VIP database | Schizophrenia strict (DSM-IV) | 39 | 32.21 ± 9.48 | 28 | PANSS POS (14.91 ± 7.14) PANSS NEG (21.40 ± 8.59) | 53 | 35.97 ± 11.32 | 56 |
| All datasets | Schizophrenia strict (DSM-IV) | 276 | 34.46 ± 11.99 | 27 | N/A | 330 | 32.36 ± 12.53 | 47 |
| VALIDATION SET PRAGUE database | First-episode psychosis | 43 | 29.18 ± 6.14 | 56 | PANSS POS (11.33 ± 3.63) PANSS NEG (13.64 ± 5.86) | 90 | 27.74 ± 6.74 | 55 |

The validation set is exclusively used for evaluation of the generalization of the learnt predictive model.

the volume of subcortical regions and the average thickness of cortical parcels.

## Machine learning algorithms

Classification analyses were first performed with linear support vector machine (SVM), implemented in the scikit-learn python library (19). However, this classifier produces dense patterns of predictors that are difficult to interpret. Although some methods exist to define thresholds to uncover brain regions that significantly contribute to the classification process (20, 21), they do not produce interpretable weight maps per se. In the context of predictive signature discovery, it is crucial to understand the brain structural patterns that underpin the prediction. We therefore seek a complementary approach able to select a reduced number of predictive regions.

Despite the fact that l1-regularized (sparse) probabilistic predictors, such as ElasticNet classifier (22), often have been advocated as leading to more interpretable models, they appear to provide scattered and unstable patterns in practice (23).

One solution to obtain more interpretable models is to exploit the known structure of brain MRI images in order to force the solution to adhere to biological priors, thereby producing more plausible and interpretable solutions. Indeed, MRI data are naturally encoded on a three-dimensional space where VBM voxels have neighbours in the three-dimensional image grid, and vertices have neighbours in the mesh of cortical surface. We here use an alternative to the SVM classifier to obtain *Structured sparsity* (i.e. a pattern of predictors that is both sparse and structured): the ElasticNet-Total Variation (Enet-TV) penalty (24) that enforces spatial smoothness of the solution while simultaneously segmenting predictive regions from the background. It has been demonstrated that this classifier generates a coherent, parsimonious and interpretable predictive map: a highly desirable characteristic in the scope of predictive signature discovery. For example, Enet-TV has been successfully used, recently, in the prediction of prehallucination functional MRI patterns in a clinical population of schizophrenia patients (25).

Therefore, grey matter VBM and vertex-based cortical thickness features' models were evaluated using SVM, Enet and Enet-TV classifiers, while ROI-based model was only conducted using SVM and Enet given that there is no explicit spatial structure in this last set of features. We expect classifiers to perform similar in terms of absolute prediction performance, but in addition the Enet-TV models will produce an interpretable predictive signature of the disorder organized in few regions of imaging features (voxels or vertices). For all analyses, we included age, gender and site as covariates.

## Cross-validation and performance assessment

Performance was evaluated using a double cross-validation (CV) scheme. It consists of two nested cross-validation loops. In the outer (external) loop, a set of subjects is considered as the training data, while the remaining subjects are held out and used as the test data. The test sets are exclusively used for model assessment, while the training sets were partitioned into sub-training and validation sets, using a nested five-fold CV, to set all regularization parameters (regularization parameter that controls the spatial smoothness of the solution, i.e., lambda_TV).

The splitting process of the samples into train and test subsets is crucial for performance evaluation. In order to investigate the reproducibility of prediction performance across sites, we chose to carry out a leave-one-site-out procedure (see Fig. 1) for the outer CV. Subjects from all sites except one are referred to as the training data, while all subjects of the remaining site are held out and used as the test data. This intersite setting is paramount in order to assess the reproducibility of a prediction model on completely independent datasets.

The classifier performances were assessed by computing the balanced accuracy, sensitivity and specificity using the test samples. Sensitivity is defined as the ability to correctly classify patients, whereas specificity evaluates the ability to identify healthy controls. The balanced accuracy score is defined as the average of the sensitivity and specificity. We also implemented the receiver operating characteristic (ROC) curve for each classifier, from which the area under the curve (AUC) was computed. To measure the significance of the prediction scores against chance level, we used an exact binomial test.

Along with the prediction performances, we also targeted a more important goal: the estimation of reproducible weight maps against variations of the learning samples. Indeed, clinicians expect that the identified signature of the disorder, that is the non-null weights of the weight map, to be similar if other patients, with similar clinical conditions, would have been used. We therefore used a similarity measure to assess the stability of those weights maps across re-sampling: the mean correlation between pairs of weights maps computed across the four-folds and denoted $r_\beta$. This measure of stability
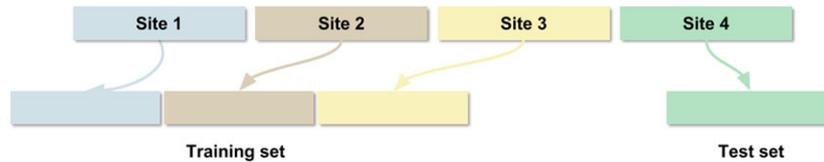
*Fig. 1.* Leave-one-site-out procedure. [Colour figure can be viewed at wileyonlinelibrary.com]

was evaluated on the weight maps provided by both sparse classifiers: Enet and Enet-TV. Indeed, SVM yields dense weight map and thus comparing the region selected across fold is not relevant.

### Interpreting the predictive signature

In order to analyse the brain regions that drive the prediction, we refitted the best model, determined by the CV, on all subjects of the database and we extracted the associated discriminative weights. These weights revealed the spatial patterns that best discriminate schizophrenia patients from healthy controls. The weights revealed the relative contribution of each feature to the decision function. Negative weights reflect that the associated features (local grey matter density or thickness of the cerebral cortex) were higher in controls than in patients with schizophrenia. Positive weights reflect the converse: feature value is higher in patients than in the controls.

### Brain signature and symptomatic level

The neuroanatomical predictive signature can be applied to each individual scan to produce a neural score of the disorder for each patient. In a post-hoc analysis, we investigated to what extent this neural score can track the symptomatic level. We leveraged the cognitive scores and symptom severity scales assessed on patients: patients's cognitive functions were evaluated using a battery of neuropsychological tests that are relevant to cognition abnormalities previously reported in schizophrenia: crystallized intelligence, working memory, episodic memory and executive functions. Those measurements were only available for a subset of 118 patients. Clinical symptom scores were evaluated through clinical rating of the symptoms dimensions: the Scale for the Assessment of Positive Symptoms (SAPS) and the Scale for the Assessment of Negative Symptoms (SANS). We evaluated the correlation between the neural score provided by the brain predictive signature and those clinical scores. To do so, we regressed each clinical score on the neural score (obtained with the brain signature), while controlling for the

effects of age and gender. A $P$-values threshold of 0.05 was considered as significant.

### Brain signature and medication/duration of illness

The impact of antipsychotic treatments on the brain anatomy has been previously reported in the literature (15, 26). This raises questions about the validity of the learned models and the predictive signature. Our concern was that patients and controls might be classified with regard to their medication status rather than their diagnosis. In order to discard the hypothesis of a confounding effect of medication on discriminative patterns, we conducted two additional analyses.

First, we trained a new classifier with a restricted set of features. Based on the literature, we masked out the regions that are known to be affected by antipsychotic drugs, such as the striatum (27, 28). We created a new predictive model using the remaining features and evaluated its performance.

Second, we took advantage of a validation cohort, constituted of 133 subjects: 90 healthy controls and 43 participants with first-episode psychosis (see Table 1). Some of those patients have taken antipsychotic medication. However, the duration of treatment is very limited (average: $2.56 \pm 5.1$ months). Thus, we assumed that the medication impacts on the brain are very limited in this cohort.

We evaluated the ability of the models learned on the full cohort, to predict diagnosis in this new, additional population.

These two complementary strategies were designed to ensure that the learned models are independent from medication and duration of illness effects and mainly rely on brain markers inherent to schizophrenia per se.

### Results

#### Prediction performances

Classification results obtained with the intersite cross-validation splitting strategy are presented in Table 2.

Table 2. Intersite prediction performances and stability using different sets of features and classifiers

| Features | Classifiers | AUC | Acc | Spe | Sen | $r_\beta$ |
|---|---|---|---|---|---|---|
| Grey matter VBM | SVM | 0.74* | 0.69* | 0.68 | 0.69 | – |
| | Enet | 0.76* | 0.71* | 0.68 | 0.73 | 0.34 |
| | Enet-TV | 0.74* | 0.68* | 0.68 | 0.68 | 0.74 |
| Vertex-based cortical thickness | SVM | 0.69* | 0.64* | 0.63 | 0.65 | – |
| | Enet | 0.60* | 0.61* | 0.61 | 0.61 | 0.09 |
| | Enet-TV | 0.70* | 0.66* | 0.60 | 0.71 | 0.76 |
| ROI-based features | SVM | 0.78* | 0.72* | 0.71 | 0.72 | – |
| | Enet | 0.74* | 0.69* | 0.69 | 0.70 | – |

Prediction accuracies: Sensitivity (Sen, recall rate of trans samples), Specificity (Spe, recall rate of off samples) and Balanced accuracy (Acc): (Sen+Spe)/2; AUC indicates area under the curve. $r_\beta$: mean correlation between pairs of weights maps computed across the four-folds. Significance notations: *: $P \leq 10^{-2}$.

The three classifiers did not differ in terms of absolute prediction performances. They were all able to significantly distinguish patients from healthy controls using all three feature sets. Grey matter VBM and ROI-based features seem to yield better predictive performance (with an AUC of 0.74 and 0.78, respectively) than vertex-based cortical thickness features (with an AUC of 0.70). The prediction performance for each site is reported in Table S2. To address the sex-specificities in schizophrenia, we also investigated whether the number of misclassified male subjects is significantly different from the number of misclassified female subjects. Across the three kinds of features, we found that the proportion of misclassified male and female is very similar: respectively, 32% of misclassified female and 33% of misclassified male.

Neuroanatomical predictive signature

We were also interested in the interpretability of the discriminative weight maps. Predictive weight maps yielded by the classifiers are presented in Fig. 2 (VBM and Vertex-based features) and Figure S3 of supplementary materials for ROIs-based features. When using the regular SVM classifier, the relevance of the obtained discriminative weight maps appears limited: it produces a dense map where all voxels/vertices contribute to the prediction. It is challenging to interpret without arbitrary thresholding. Understanding the structural brain patterns that drive the prediction is crucial. Meanwhile, the predictive maps obtained with TV-Enet classifier appear much more interpretable, since it provides a smooth map made of several clearly identifiable regions.

Besides the prediction scores, we also targeted a more important goal with the classifier Enet-TV: the estimation of reproducible weight maps across folds: for VBM features, the mean correlation $r_\beta = 0.74$, and for vertex-based features, the mean correlation $r_\beta = 0.76$. The weight maps yielded for each fold by Enet-TV are presented in Figures S4 and S5.

Brain signature and symptomatic level

Since the VBM feature yields better predictive performance than the vertex-based features, we restricted the correlation analysis with clinical scores to the VBM predictive signature. We found significant positive correlations between the VBM predictive signature and both, the negative symptoms scores ($r = 0.17$, $P < 3.5e^{-2}$) and the positive symptoms scores ($r = 0.18$, $P < 2.2e^{-2}$) (Table S6 and Figure S7). The predictive signature also correlated with the extent of cognitive deficits in most of the domains tested: crystallized intelligence, working memory and episodic memory.

Brain signature and medication/duration of illness influence

The classifiers we developed possibly rely more on the effects of treatment or evolution of the disease on the brain rather than on markers of the disorder to distinguish healthy controls from schizophrenia patients. To discard this hypothesis, we ran additional predictive models by masking out the regions that are known to be affected by antipsychotic medications (such as the striatum region). Even without these regions, results are encouraging as we obtained similar prediction accuracy than with the full model (see Table S8).

We also assessed the prediction performance of the learned models on an independent set of subjects with a first-episode psychosis ('validation cohort', see Table 1). This sample was not included in the learning datasets. The prediction performances obtained from those patients are presented in Table 3.

The prediction performances are promising, ranging from 64% to 76% of accuracy, depending on the features used to build the model.

Discussion

In this large intersite study, we showed that machine learning classifiers based on neuroanatomical features are able to accurately distinguish controls from schizophrenia patients in an intersite setting. A predictive neuroanatomical signature associated with the classification process can be extracted and interpreted. Moreover, the models were found independent to duration of illness and have the ability to generalize to the prediction of first-episode psychosis.
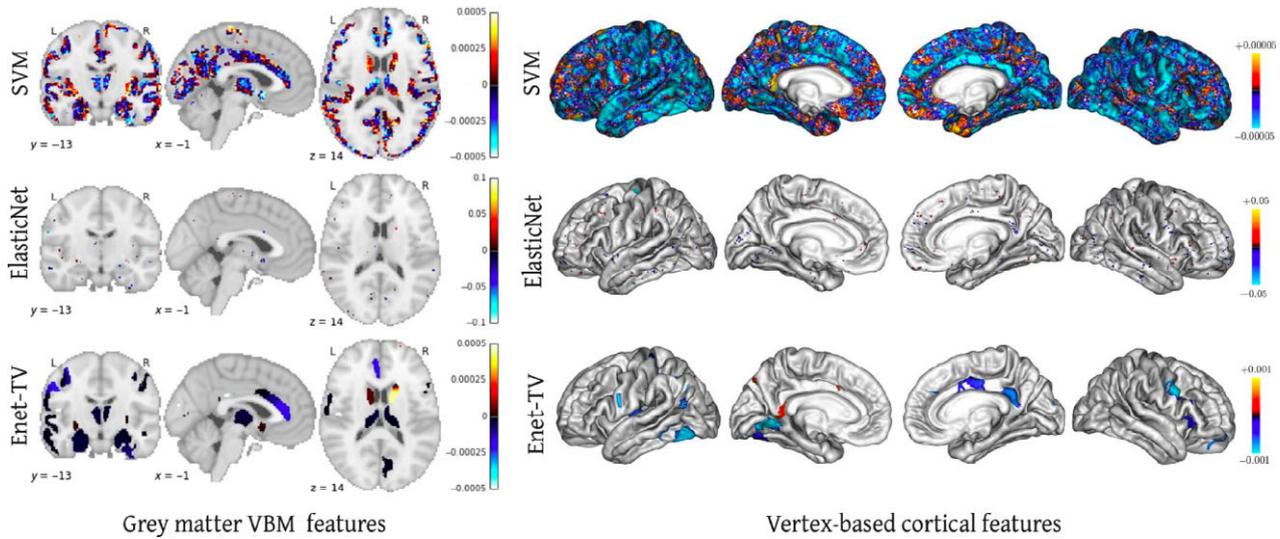
*Fig. 2.* Predictive signatures obtained with the classifiers – Top row: SVM, Middle row: Enet and Bottom row: Enet-TV. [Colour figure can be viewed at wileyonlinelibrary.com]

Table 3. Intersite prediction performances on independent subjects with first-episode psychosis. Prediction accuracies: Sensitivity (Sen, recall rate of trans samples), Specificity (Spe, recall rate of off samples) and Balanced accuracy (Acc): (Sen+Spe)/2; AUC indicates area under the curve

| Features | Classifiers | AUC | Acc | Spe | Sen |
|---|---|---|---|---|---|
| Grey matter VBM | SVM | 0.78* | 0.71* | 0.61 | 0.81 |
| | Enet-TV | 0.76* | 0.73* | 0.66 | 0.81 |
| Vertex-based cortical thickness | SVM | 0.68* | 0.64* | 0.59 | 0.69 |
| | Enet-TV | 0.64* | 0.61* | 0.52 | 0.69 |
| ROI-based volume | SVM | 0.72* | 0.66* | 0.63 | 0.69 |

Significance notations: *: $P \leq 10^{-2}$.

Prediction performances

The predictive models obtained robust intersite prediction performances that are consistent with the average prediction scores reported in the literature (7, 12). This suggests that the predictive models of schizophrenia we developed here are able to generalize to subjects from unseen sites. This is promising in the scope of cross-site classification of individuals acquired with different scanner brands, models and MRI sequences. Despite such intersite variability, a good leave-one-site-out accuracy is achieved. Furthermore, the use of more sophisticated methods to harmonize the MRI scans directly at acquisition or during preprocessing steps (29) could potentially improve this absolute prediction performance. It constitutes a challenge for future clinical applications.

Interestingly, the highest predictive performances are delivered with VBM features in this current study. However, lately, cortical features are becoming more popular than VBM features because the cortical measurements (such as thickness) have a well-defined neurobiological interpretation. Conversely, VBM provide local changes of volumes in a template space through a nonlinear registration. This always raises question whether the observed changes stem from a true variation of tissue volume in the subject space or if they are the result of the normalization or both. Nevertheless, this study is consistent with a recent study (16) and tends to demonstrate that VBM features capture more predictive information than cortical features. This study assesses the efficacy of sMRI data for diagnostic prediction in psychosis by evaluating the discriminative power of a wide range of commonly used sMRI features including grey matter voxel-based morphometry (VBM), vertex-based cortical thickness and region of interest volumetric measures. Results show that grey matter VBM (without data reduction) are the feature types leading to highest accuracies. The poorer performance of vertex-based cortical features may be related to their restricted spatial extent, which excludes all subcortical structures. Indeed, major structural abnormalities in schizophrenia have been described in those regions of high topological complexity. Grey matter density maps are likely reflecting relevant brain morphological variance that is not modelled by the surface-based cortical thickness features (30).

Besides the absolute prediction performance, we are also interested in the identification of a neuroanatomical predictive signature of schizophrenia.

Neuroanatomical predictive signature

Identifying a neuroanatomical signature of schizophrenia that is clinically interpretable is crucial. All things considered, while the state-of-the-art SVM classifier provides dense patterns of predictors that are clinically uninterpretable, the Enet classifier yields predictive patterns that are sparse and scattered across the brain. Using the advanced machine learning (Enet-TV) classifier (which performs similar than the SVM and Enet, in terms of absolute prediction performances), we were able to identify an interpretable neuroanatomical predictive signature of schizophrenia, that is organized in, brain regions that are in line with the literature. Moreover, the predictive signature yielded by Enet-TV is reproducible across folds, with similar predictors selected when different samples are used in the training phase.

The predictive signature yielded is consistent across the three types of features (identified regions from (i) whole brain voxels or (ii) cortical vertices and (iii) atlas-based ROIs) (see Table S9 and S10 for details of this signature). The identified patterns appear largely consistent with available neural data in schizophrenia and may fill the criteria to become a signature of the disorder. We indeed found that classification of patients with schizophrenia relied on reduced grey matter compared to healthy controls in the cingulate gyrus, precentral and postcentral gyrus, temporal pole, hippocampus, amygdala and thalamus. These regional deficits of grey matter in schizophrenia patients have been consistently reported in univariate studies (31–36).

On the other hand, we found a regional increase of grey matter in schizophrenia patients compared to healthy controls in the putamen, caudate and pallidum. These local increases of GM in schizophrenia were also frequently reported in previous studies (31–37).

The predictive signature is also consistent with discriminative regions identified in previous machine learning studies. For example, in a recent study (38) with a large population of 326 participants of drug-naïve first-episode psychosis (FEP) patients and demographically matched healthy controls, the regions contributing to the classification mainly included the left inferior parietal, left rostral anterior cingulate, left rostral middle frontal, right caudal middle frontal, right inferior parietal, right lingual and right temporal pole cortex. Findings in this sample of never-treated FES patients are related to the regions identified in this current study. It suggests that the found predictive signature of the disorder could potentially be used for identifying structural brain alterations in individuals with schizophrenia.

Furthermore, significant correlations were found between this predictive signature and both negative and positive symptom scores. Such a result is consistent with the literature where negative symptoms have already been reported to be associated with the extent of structural brain abnormalities in schizophrenia (7, 39). Additionally, the neural score obtained from the predictive signature is also correlated with the extent of cognitive impairments in all cognitive domains that are known to be impacted in schizophrenia, except executive functions. This result is promising since it paves the way towards the use of a neuroanatomical signature as an objective measure to monitor the evolution of the disorder.

Medication/duration of illness influence

We also tested for the prediction performances obtained on the first-episode psychosis cohort. Indeed, from a clinical perspective, the true value of MRI-based prediction lies in the early diagnosis. Indeed, accurately predicting chronic schizophrenia patients affected by the disease for a long time does not provide ground-breaking insight. Instead, what is clinically relevant is the identification of patients who are still at an early stage of the disease.

Interestingly, our predictive models appear able to accurately classify first-episode psychosis subjects as patients. This finding suggests that these classifiers mainly rely on true markers of schizophrenia rather than medication effects or duration of illness. The identified neuroanatomical predictive signature seems to generalize to the detection of patients at the early stage of the disorder. Furthermore, because providing early care to reduce the duration of untreated psychosis has been identified as a predictor of long-term outcome in schizophrenia (40), present findings directly question the systematic use of sMRI combined with predictive models to assist clinicians in the early stages of the disorder.

Future work

We demonstrated in this study that it is possible to accurately discriminate schizophrenia patients from controls, using structural MRI. At this stage, this does not imply that such models are able to distinguish patients with various psychiatric conditions. In order to demonstrate the clinical relevance of predictive models such as the one

developed in this study, the next step would be to evaluate the specificity of the classifiers in differential diagnosis situations. There is now an urgent need for transdiagnostic studies able to compare the specificity of the identified neuroanatomical predictive signature in schizophrenia but also in bipolar disorder or autism spectrum disorder.

These results highlight the existence of a neuroanatomical signature of schizophrenia, shared by a majority of patients across different sites and already present at the early stage of the disorder. Moreover, this signature is associated with the symptoms severity and the amount of cognitive deficit. Such neuroanatomical signature is made publicly available at ftp://ftp.cea.fr//pub/unati/brainomics/papers/scz_predict_vbm.

## Declaration of interest

None.

## References

1. van Erp TGM, Hibar DP, Rasmussen JM et al. Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. Mol Psychiatry 2016;**21**:547–553.

2. Sun D, van Erp TGM, Thompson PM et al. Elucidating a magnetic resonance imaging-based neuroanatomic biomarker for psychosis: classification analysis using probabilistic brain atlas and machine learning algorithms. Biol Psychiatry 2009;**66**:1055–1060.

3. Orrù G, Pettersson-Yeo W, Marquand AF, Sartori G, Mechelli A. Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. Neurosci Biobehav Rev 2012;**36**:1140–1152.

4. Kambeitz J, Kambeitz-Ilankovic L, Leucht S et al. Detecting neuroimaging biomarkers for schizophrenia: a meta-analysis of multivariate pattern recognition studies. Neuropsychopharmacology 2015;**40**:1742–1751.

5. Lu X, Yang Y, Wu F et al. Discriminative analysis of schizophrenia using support vector machine and recursive feature elimination on structural MRI images. Medicine 2016;**95**:e3973.

6. Sabuncu MR, Konukoglu E. Alzheimer's Disease Neuroimaging Initiative. Clinical prediction from structural brain MRI scans: a large-scale empirical study. Neuroinformatics 2015;**13**:31–46.

7. Rozycki M, Satterthwaite TD, Koutsouleris N et al. Multisite machine learning analysis provides a robust structural imaging signature of schizophrenia detectable across diverse patient populations and within individuals. Schizophr Bull 2017;**44**:1035–1044.

8. Vita A, De Peri L, Silenzi C, Dieci M. Brain morphology in first-episode schizophrenia: a meta-analysis of quantitative magnetic resonance imaging studies. Schizophr Res 2006;**82**:75–88.

9. Adriano F, Spoletini I, Caltagirone C, Spalletta G. Updated meta-analyses reveal thalamus volume reduction in patients with first-episode and chronic schizophrenia. Schizophr Res 2010;**123**:1–14.

10. Shepherd AM, Laurens KR, Matheson SL, Carr VJ, Green MJ. Systematic meta-review and quality assessment of the structural brain alterations in schizophrenia. Neurosci Biobehav Rev 2012;**36**:1342–1356.

11. Orban P, Dansereau C, Desbois L et al. Multisite generalizability of schizophrenia diagnosis classification based on functional brain connectivity. Schizophr Res 2018;**192**:167–171.

12. Nieuwenhuis M, van Haren NEM, Hulshoff Pol HE, Cahn W, Kahn RS, Schnack HG. Classification of schizophrenia patients and healthy controls from structural MRI scans in two large independent samples. NeuroImage 2012;**61**:606–612.

13. Rozycki M, Satterthwaite TD, Koutsouleris N et al. Multisite machine learning analysis provides a robust structural imaging signature of schizophrenia detectable across diverse patient populations and within individuals. Schizophr Bull 2018;**44**:1035–1044.

14. Kawasaki Y, Suzuki M, Kherif F et al. Multivariate voxel-based morphometry successfully differentiates schizophrenia patients from healthy controls. NeuroImage 2007;**34**:235–242.

15. Radua J, Borgwardt S, Crescini A et al. Multimodal meta-analysis of structural and functional brain changes in first episode psychosis and the effects of antipsychotic medication. Neurosci Biobehav Rev 2012;**36**:2325–2333.

16. Salvador R, Radua J, Canales-Rodríguez EJ et al. Evaluation of machine learning algorithms and structural features for optimal MRI-based diagnostic prediction in psychosis. PLoS ONE 2017;**12**:e0175683.

17. Ashburner J. Computational anatomy with the SPM software. Magn Reson Imaging 2009;**27**:1163–1174.

18. Ashburner J, Friston KJ. Unified segmentation. NeuroImage 2005;**26**:839–851.

19. Pedregosa F, Varoquaux G, Gramfort A et al. Scikit-learn: machine Learning in Python. J Mach Learn Res 2011;**12**:2825–2830.

20. Wang Z, Childress AR, Wang J, Detre JA. Support vector machine learning-based fMRI data group analysis. NeuroImage 2007;**36**:1139–1151.

21. Gaonkar B, Davatzikos C. Analytic estimation of statistical significance maps for support vector machine based multi-variate image analysis and classification. NeuroImage 2013;**78**:270–283.

22. Zou H, Hastie T. Regularization and variable selection via the elastic net. J Royal Stat Soc 2005;**67**:301–320.

23. Grosenick L, Klingenberg B, Katovich K, Knutson B, Taylor JE. Interpretable whole-brain prediction analysis with GraphNet. NeuroImage 2013;**72**:304–321.

24. Hadj-Selem F, Löfstedt T, Dohmatob E et al. Continuation of Nesterov's smoothing for regression with structured sparsity in high-dimensional neuroimaging. IEEE Trans Med Imaging 2018:1–1.

25. de Pierrefeu A, Fovet T, Hadj-Selem F et al. Prediction of activation patterns preceding hallucinations in patients with schizophrenia using machine learning with structured sparsity. Hum Brain Mapp 2018;**39**:1777–88.

26. Roiz-Santiañez R, Suarez-Pinilla P, Crespo-Facorro B. Brain structural effects of antipsychotic treatment in schizophrenia: a systematic review. Curr Neuropharmacol 2015;**13**:422–434.

27. Smieskova R, Fusar-Poli P, Allen P et al. The effects of antipsychotics on the brain: what have we learnt from structural imaging of schizophrenia?–A systematic review. Curr Pharm Des 2009;**15**:2535–2549.

28. Torres US, Portela-Oliveira E, Borgwardt S, Busatto GF. Structural brain changes associated with antipsychotic treatment in schizophrenia as revealed by voxel-based morphometric MRI: an activation likelihood estimation meta-analysis. BMC Psychiatry 2013;**13**:342.

29. Fortin J-P, Sweeney EM, Muschelli J, Crainiceanu CM, Shinohara RT. Alzheimer's Disease Neuroimaging Initiative. Removing inter-subject technical variability in magnetic resonance imaging studies. NeuroImage 2016;**132**:198–212.

30. Hutton C, Draganski B, Ashburner J, Weiskopf N. A comparison between voxel-based cortical thickness and voxel-based morphometry in normal aging. NeuroImage 2009;**48**:371–380.

31. Honea R, Crow TJ, Passingham D, Mackay CE. Regional deficits in brain volume in schizophrenia: a meta-analysis of voxel-based morphometry studies. Am J Psychiatry 2005;**162**:2233–2245.

32. Glahn DC, Laird AR, Ellison-Wright I et al. Meta-analysis of gray matter anomalies in schizophrenia: application of anatomic likelihood estimation and network analysis. Biol Psychiatry 2008;**64**:774–781.

33. Fornito A, Yücel M, Patti J, Wood SJ, Pantelis C. Mapping grey matter reductions in schizophrenia: an anatomical likelihood estimation analysis of voxel-based morphometry studies. Schizophr Res 2009;**108**:104–113.

34. Torres US, Duran FLS, Schaufelberger MS et al. Patterns of regional gray matter loss at different stages of schizophrenia: a multisite, cross-sectional VBM study in first-episode and chronic illness. Neuroimage Clin 2016;**12**:1–15.

35. Kim G-W, Kim Y-H, Jeong G-W. Whole brain volume changes and its correlation with clinical symptom severity in patients with schizophrenia: a DARTEL-based VBM study. PLoS ONE 2017;**12**:e0177251.

36. Shenton ME, Dickey CC, Frumin M, McCarley RW. A review of MRI findings in schizophrenia. Schizophr Res 2001;**49**:1–52.

37. Hulshoff Pol HE, Schnack HG, Mandl RC et al. Focal gray matter density changes in schizophrenia. Arch Gen Psychiatry 2001;**58**:1118–1125.

38. Xiao Y, Yan Z, Zhao Y et al. Support vector machine-based classification of first episode drug-naïve schizophrenia patients and healthy controls using structural MRI. Schizophr Res 2017. https://doi.org/10.1016/j.schres.2017.11.037

39. Ren W, Lui S, Deng W et al. Anatomical and functional brain abnormalities in drug-naive first-episode schizophrenia. Am J Psychiatry 2013;**170**:1308–1316.

40. Penttilä M, Jääskeläinen E, Hirvonen N, Isohanni M, Miettunen J. Duration of untreated psychosis as predictor of long-term outcome in schizophrenia: systematic review and meta-analysis. Br J Psychiatry 2014;**205**:88–94.

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Figure S1.** ROI-based features: Discriminative weights associated to each feature.

**Figure S2.** VBM discriminative weight maps yielded at each fold.

**Figure S3.** Vertex-based discriminative weight maps yielded at each fold.

**Figure S4.** Correlation between the neuroanatomical signature score and the negative and positive symptoms scores (SANS and SAPS) of patients.

**Table S1.** Intersite prediction performances where features known to be affected by antipsychotic drugs have been removed from the analysis.

**Table S2.** MRI Acquisition parameters details.

**Table S3.** Accuracy of prediction on each independent site.

**Table S4.** Associations between cognitive and symptoms severity scores and the predictive signature.

**Table S5.** Grey matter voxel-based morphometry features: Discriminative clusters. Clusters are presented ordered by weight.

**Table S6.** Vertex-based cortical thickness features: Discriminative clusters. Clusters are presented ordered by weight.