



Interpretable machine learning model for characterizing magnetic susceptibility-based biomarkers in first episode psychosis

Pamela Franco ^a, Cristian Montalba ^{b,c,d}, Raúl Caulier-Cisterna ^e, Carlos Milovic ^f, Alfonso González ^{g,h}, Juan Pablo Ramirez-Mahaluf ^g, Juan Undurraga ⁱ, Rodrigo Salas ^{b,j}, Nicolás Crossley ^{b,c,g}, Cristian Tejos ^{b,c,k}, Sergio Uribe ^{l,*}

^a Energy Transformation Center, Faculty of Engineering, Universidad Andrés Bello, Santiago, Chile

^b Millennium Institute for Intelligent Healthcare Engineering - iHEALTH, Santiago, Chile

^c Biomedical Imaging Center, Pontificia Universidad Católica de Chile, Santiago, Chile

^d Radiology Department, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile

^e Department of Informatics and Computing, Faculty of Engineering, Universidad Tecnológica Metropolitana, Santiago, Chile

^f School of Electrical Engineering, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

^g Department of Psychiatry, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile

^h School of Medicine, Universidad Finis Terrae, Chile

ⁱ Department of Neurology and Psychiatry, Faculty of Medicine, Clínica Alemana Universidad del Desarrollo, Santiago, Chile

^j Biomedical Engineering School, Faculty of Engineering, Universidad de Valparaíso, Valparaíso, Chile

^k Department of Electrical Engineering, School of Engineering, Pontificia Universidad Católica de Chile, Santiago, Chile

^l Department of Medical Imaging and Radiation Sciences, School of Primary and Allied Health Care, Faculty of Medicine, Nursing and Health Sciences, Monash University, Australia

ARTICLE INFO

Keywords:

Interpretable machine learning
Magnetic resonance imaging
Biomarkers, artificial intelligence for medical imaging
Quantitative susceptibility mapping

ABSTRACT

Background and Purpose: Several studies have shown changes in neurochemicals within the deep-brain nuclei of patients with psychosis. These alterations indicate a dysfunction in dopamine within subcortical regions affected by fluctuations in iron concentrations. Quantitative Susceptibility Mapping (QSM) is a method employed to measure iron concentration, offering a potential means to identify dopamine dysfunction in these subcortical areas. This study employed a random forest algorithm to predict susceptibility features of the First-Episode Psychosis (FEP) and the response to antipsychotics using Shapley Additivity Explanation (SHAP) values.

Methods: 3D multi-echo Gradient Echo (GRE) and T1-weighted GRE were obtained in 61 healthy-volunteers (HV) and 76 FEP patients (32 % Treatment-Resistant Schizophrenia (TRS) and 68 % treatment-Responsive Schizophrenia (RS)) using a 3T Philips Ingenia MRI scanner. QSM and R2* were reconstructed and averaged in twenty-two segmented regions of interest. We used a Sequential Forward Selection as a feature selection algorithm and a Random Forest as a model to predict FEP patients and their response to antipsychotics. We further applied the SHAP framework to identify informative features and their interpretations. Finally, multiple correlation patterns from magnetic susceptibility parameters were extracted using hierarchical clustering.

Results: Our approach accurately classifies HV and FEP patients with 76.48 ± 10.73 % accuracy (using four features) and TRS vs RS patients with 76.43 ± 12.57 % accuracy (using four features), using 10-fold stratified cross-validation. The SHAP analyses indicated the top four nonlinear relationships between the selected features. Hierarchical clustering revealed two groups of correlated features for each study.

Conclusions: Early prediction of treatment response enables tailored strategies for FEP patients with treatment resistance, ensuring timely and effective interventions.

1. Introduction

Psychosis can occur in various conditions, including schizophrenia,

and affective disorders [1]. Some individuals may have a single episode of psychosis, while others may develop schizophrenia or other psychotic or mood disorders [2]. Early intervention and appropriate treatment

* Corresponding author.

E-mail address: sergio.uribe@monash.edu (S. Uribe).

<https://doi.org/10.1016/j.cmpb.2025.109067>

Received 29 January 2025; Received in revised form 2 September 2025; Accepted 3 September 2025

Available online 6 September 2025

0169-2607/© 2025 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

during FEP, including antipsychotic medications, can play a crucial role in managing symptoms and potentially preventing the progression to more severe psychotic disorders, such as schizophrenia [3]. Despite an initial treatment, some individuals may develop schizophrenia, and a subset of them (about one-third) may develop treatment-resistant (TSR) symptoms [4]. Clozapine is the treatment of choice for TSR. However, there is hesitancy to prescribe clozapine due to its side effects and its position as a "treatment of last resort" [5]. Consequently, there is a significant delay in its initiation [6] and a significant underutilization worldwide [7] despite its effectiveness in managing symptoms and increasing survival rates.

The classical hypothesis of psychosis posits that the cause of this disease is dopamine hyperactivity in the mesolimbic dopamine pathway at dopamine D2 receptors [8]. A recent theory suggests that dopamine hyperactivity is a downstream consequence of glutamate dysregulation in the prefrontal cortex, resulting in hypofunctional NMDA receptors on GABA interneurons [9]. Regardless of the underlying mechanism of dopamine hyperactivity, this theory is one of the most influential in the etiology of schizophrenia [9]. These interconnected pathological processes share a common element: iron [10]. Iron is a co-factor in neurotransmitter biosynthesis and is associated with neuroinflammation that could be produced in functionally abnormal Grey Matter (GM) nuclei [11].

Histological studies in schizophrenia have shown some associated brain abnormalities, including GM loss [12], compromised white matter connectivity [13], and aberrant functional connectivity [14]. Some authors reported higher iron content in the caudate nucleus [15] and globus pallidus [16], whereas another study found no changes in the caudate nucleus [17]. Recent neuroimaging developments have allowed in vivo quantification of iron concentrations in the brain [18]. Quantitative Susceptibility Mapping (QSM), a Magnetic Resonance Imaging (MRI) technique, enables measuring iron concentrations using its paramagnetic characteristics, providing evidence of iron concentration changes in brain nuclei [19] that are associated with schizophrenia or psychosis. For instance, studies reported a strong linear correlation between iron concentration and bulk magnetic susceptibility in GM structures [20].

Some authors demonstrated the possibility of refining disease monitoring by considering a limited number of dopamine pathways (nigrostriatal and tuberoinfundibular pathways) retrieved from QSM images [11]. In particular, Xu et al. [11] and Garcia et al. [21] used a statistical linear mixed model to identify abnormal iron concentration in deep GM nuclei in FEP patients and explain the relationship between QSM and the apparent transverse relaxation rate ($R2^*$) values and a disease severity score.

Compared to conventional statistical models, Machine Learning (ML) algorithms achieve high predictive accuracy and can reveal intricate data patterns, without strictly adhering to underlying statistical assumptions [22]. ML operates on statistical learning theory, offering superior predictive capabilities, albeit with a more intricate understanding and explanation process [23]. ML techniques are complementary tools to inferential statistics, enabling the exploration of nonlinear relationships among variables and accommodating vast amounts of high-dimensional data [24]. Utilizing ML techniques presents an alternative approach for leveraging QSM and $R2^*$ data to comprehend the dopamine pathways in psychosis. These techniques would enable us to develop a classifier to distinguish between FEP and HV and would also allow to identify treatment resistant subjects.

MRI data analysis with ML algorithms has shown promising results in stratifying patients with brain disorders, such as chronic schizophrenia [25]. Pattern recognition, i.e., the automatic discovery of regularities in data through ML, has been applied to FEP patients [26]. This technique allows for inferences about the subjects' health status at an individual level and, thus, is well-suited for clinical decision-making.

Assessing the impact of susceptibility changes within multiple deep brain structures in psychosis is difficult due to extensive data. Different

Feature Selection (FS) methods can reduce data dimensionality and redundancy, allowing for the determination of which features discriminate best between classes. Given the increasing capacity to extract a large amount of data from images, FS methods have become essential to achieve effective classification. They provide simpler models with equivalent or even better performance, where generalization capacity may be improved. The weakness of most ML methods is the difficulty in explaining how they perform associations [27]. To deal with this issue, Shapley Additive exPlanations (SHAP) values assign each feature an importance weight for a particular prediction [23]. Its novel components include identifying feature importance measures that explain the behavior of tree models and ensembling trees as Random Forest (RF) [28].

The present study aims to pinpoint potential predictive biomarkers derived from QSM and $R2^*$ for individuals experiencing FEP, along with their response to antipsychotic treatment. To accomplish this objective, we have introduced an RF model to predict both the occurrence of FEP and the efficacy of antipsychotic medications. Furthermore, our study aims to enhance transparency in the model's decision-making process by employing the SHAP methodology to explain prediction results. By leveraging SHAP values, our research seeks to improve the interpretability of predictive biomarkers, facilitate the identification of relevant features, provide insights into biomarker relationships, enable individualized predictions, and validate predictive models within our cohort.

Therefore, the technical novelty of this study lies in its innovative integration of advanced neuroimaging techniques, particularly QSM and $R2^*$ values, with ML methods to predict FEP and response to antipsychotic treatments. A significant contribution is introducing an RF model incorporating these neuroimaging biomarkers to forecast treatment outcomes in FEP patients. While machine learning has been applied to neuroimaging in other contexts, this study is distinct in leveraging RF models to identify patterns in QSM and $R2^*$ data, specifically for early psychosis prediction and treatment response.

An additional innovative aspect of this study is the use of SHAP, which enhances the interpretability of the RF model. By explaining the importance of individual features in model predictions, SHAP addresses a common challenge in machine learning, understanding how predictions are made. This transparency is crucial, enabling clinicians and researchers to understand the relationship between neuroimaging biomarkers and clinical outcomes. Furthermore, the study employs FS techniques to reduce dataset complexity, ensuring that the RF model remains efficient without sacrificing predictive power. This approach reduces the computational burden and improves model generalization, providing robust and applicable predictions across different datasets.

2. Methods

2.1. Study population

We used the same data from a previous prospective study [21]. We enrolled seventy-six patients who were diagnosed with a first psychotic disorder according to the Mini International Neuropsychiatric Interview (MINI) [29]. Upon their initial presentation to mental health services, these patients received treatment according to local guidelines, which included antipsychotic medication ("*Guías Clínicas AUGÉ Tratamiento de Personas Desde El Primer Episodio de Esquizofrenia, 2017*"). Patients were recruited at the Instituto Psiquiátrico Dr. J. Horwitz in Santiago, Chile. They did not have any neurological comorbidity related to psychiatric disorders or psychomotor agitation. The inclusion and exclusion criteria were assessed by their treating psychiatrists and through standardized evaluations at the time of an initial MRI scanning. The history of substance consumption was not considered an exclusion criterion. On average, patients received antipsychotic treatment for 36 ± 46.5 days before the baseline MRI scan. This variable was recorded and considered a potential confounder, as variability in treatment duration before scanning may influence QSM and $R2^*$ measures. However, given the

heterogeneity and limited sample size, we did not adjust for treatment duration in the machine learning models.

Additionally, we recruited sixty-one healthy volunteers (HV) considering as exclusion criteria any history of psychotic disorders, first-degree family with history of psychotic disorders, and a record of any mental illness.

The study was approved by the Ethics Committee of the Pontificia Universidad Católica de Chile (Ref: 15-297) and Servicio de Salud Metropolitano Norte. Written informed consent was obtained from all participants.

2.2. Clinical assessment

The cohort was assessed clinically using the Positive and Negative Syndrome Scale (PANSS) [30] and a neuropsychological assessment [31, 32]. Doses of antipsychotics were converted to chlorpromazine equivalents using the Defined Daily Dose method [33].

All seventy-six FEP patients were re-evaluated to classify as Treatment-Resistant Schizophrenia (TRS) or Treatment-Responsive Schizophrenia (RS) according to the fulfilled TRRIP criteria for treatment resistance [34].

2.3. Magnetic resonance neuro images acquisition

Multi-echo Gradient Echo (GRE) and T1-weighted GRE sequences were acquired in a 3T Philips Ingenia MRI scanner with an 8-channel SENSE head coil for all subjects. 3D multi-echo GRE acquisition parameters were: voxel size of $0.59 \times 0.59 \times 1 \text{ mm}^3$, Repetition Time (TR) of 4 ms, Echo Time (TE) of 7.2 ms, inter-echo spacings (ΔTE) of 6.2 ms, 5 echoes, bandwidth of 550.5 Hz, matrix size of $352 \times 352 \times 160$, Flip Angle (FA) of 17° , and total acquisition time of 6.8 min. 3D T1-weighted GRE were acquired with the following parameters: voxel size of $0.5 \times 0.5 \times 1 \text{ mm}^3$, TR of 7.75 ms, TE of 3.57 ms, bandwidth of 191 Hz, matrix size of $480 \times 480 \times 341$, and FA of 8° .

2.4. Quantitative susceptibility and $R2^*$ mapping

QSM and $R2^*$ maps were reconstructed from phase and magnitude 3D multi-echo GRE images, respectively. The QSM reconstruction consisted of a four-step pipeline. First, we applied the Rapid Open-Source Minimum spanning tree algorithm (ROMEO) to unwrap the phase [35]. Second, we used a linear fit to extract and remove the phase offset from the multi-echo phase data. Echoes were then combined using a weighted average to get the total field map. Weights were calculated as the multiplication of the magnitude data at each echo by its echo time, as this optimizes the contrast-to-noise ratio in the phase data, matching the $R2^*$ of the tissues. This procedure was performed to account for the acquisitions' relatively low Signal-to-Noise Ratio (SNR). Third, we removed the background components from the total field estimation, mostly from air-tissue and air-bone interfaces, using the Projection onto Dipole Fields (PDF) method [36]. Finally, we obtained the susceptibility distribution from the local field map (obtained in the previous step) using the Weak Harmonics QSM (WH-QSM) algorithm with a Total Variation (TV) regularization [37], available in the FANSI Nonlinear Susceptibility Inversion (FANSI) toolbox ([38] <http://gitlab.com/cmilovic/FANSI-toolbox>). The TV regularization weight $\alpha = 10^{-3}$ was optimized heuristically using an L-curve analysis [39]. Other parameters were fixed following the literature: harmonic consistency weight $\mu_h = 0.05$, 3500 iterations with stopping convergence parameter of 10^{-3} , and harmonic constraint weight $\beta = 10^6$. We reconstructed the $R2^*$ maps using a linear fit of the logarithm of the input value, method which is available in the FANSI toolbox.

To obtain the deep GM and subcortical brain nuclei areas that represent the dopamine pathways, we used the Multicontrast PD25 Atlas (version 2019). This contains twenty-two ROIs: left and right red nucleus, substantia nigra, subthalamic nucleus, caudate, putamen, globus

pallidus externa, and interna, thalamus, hippocampus, nucleus accumbens, and amygdala. Since the twenty-two ROIs from the atlas were in ICBM152 space [40–42], we normalized the images bringing them to the Montreal Neurological Institute (MNI) space of dimensions $91 \times 109 \times 91$ (voxel size: $2 \times 2 \times 2 \text{ mm}^3$). The magnitude of the multi-echo GRE was then co-registered with the T1-weighted image. The T1-weighted GRE was normalized, obtaining the deformation field. We applied this deformation field to the QSM and $R2^*$ maps. Finally, we computed each subject's mean QSM and $R2^*$ values from each labeled ROI, obtaining forty-four features for each subject's map. All these pre-processing steps were performed using SPM12 (Wellcome Centre for Human Neuroimaging, University College London, London, UK, www.fil.ion.ucl.ac.uk/spm) [43], as shown in Fig. 1.

2.5. QSM and $R2^*$ parameters analysis using the machine learning approach

Two ML models were designed to predict those regions whose QSM and $R2^*$ values adequately discriminate between different groups of psychosis patients. One of these ML models helps differentiate between HV and those who have experienced a FEP (HV vs. FEP patients). The other ML model helps determine whether patients respond to a specific treatment for their disorder (RS vs. TRS patients). These ML strategies were developed using Python 3.11.

We employed a comprehensive ML strategy to develop a robust predictive model that combined several techniques, including Sequential Forward Selection (SFS), RF, hyperparameter grid search, cross-validation, and SHAP values.

2.5.1. Feature selection

We initially used Ridge regression, a linear model with L2 regularization, to assess feature importance across all available features. The coefficients from the trained model indicated the relative importance of each feature. To refine the feature set, we employed SFS. SFS was implemented as a wrapped-based method using cross-validation to iteratively add features that maximize model performance. Ridge regression was only used for initial feature ranking and not for the final feature set selection.

2.5.2. Model training

We trained an RF classifier using the selected features from the SFS process. RF is an ensemble learning method that combines multiple decision trees to improve predictive performance and reduce overfitting. We generated a decision surface plot for the RF classifier to visualize the classification boundaries. The decision surface represents the regions of the feature space where the classifier assigns different class labels. Given the high-dimensional nature of the data, we reduced it to two principal components for visualization. The resulting decision surface provides insight into how the RF classifier separates classes based on the selected features. The ensemble approach of RF results in more robust and smoother decision boundaries than individual decision trees, capturing complex patterns and interactions within the data.

2.5.3. Hyperparameter optimization

To optimize the performance of the RF model, we conducted a hyperparameter grid search using stratified cross-validation. This involved systematically searching through a predefined grid of hyperparameters and selecting the combination that maximizes the model performance. General parameters were set and tested, including the number of estimations (decision trees) being twofold from 10 to 500, maximum number of features (consider how many features on split each node): sqrt, log2, and none; maximum depth (the allowed number of splits by each decision trees) between 1 and 20, minimum samples split (specifies the minimum number of samples required to split an internal node) between 5 and 50, and maximum leaf nodes (the maximum number of leaf nodes a decision tree can have) between 2 and 10. These

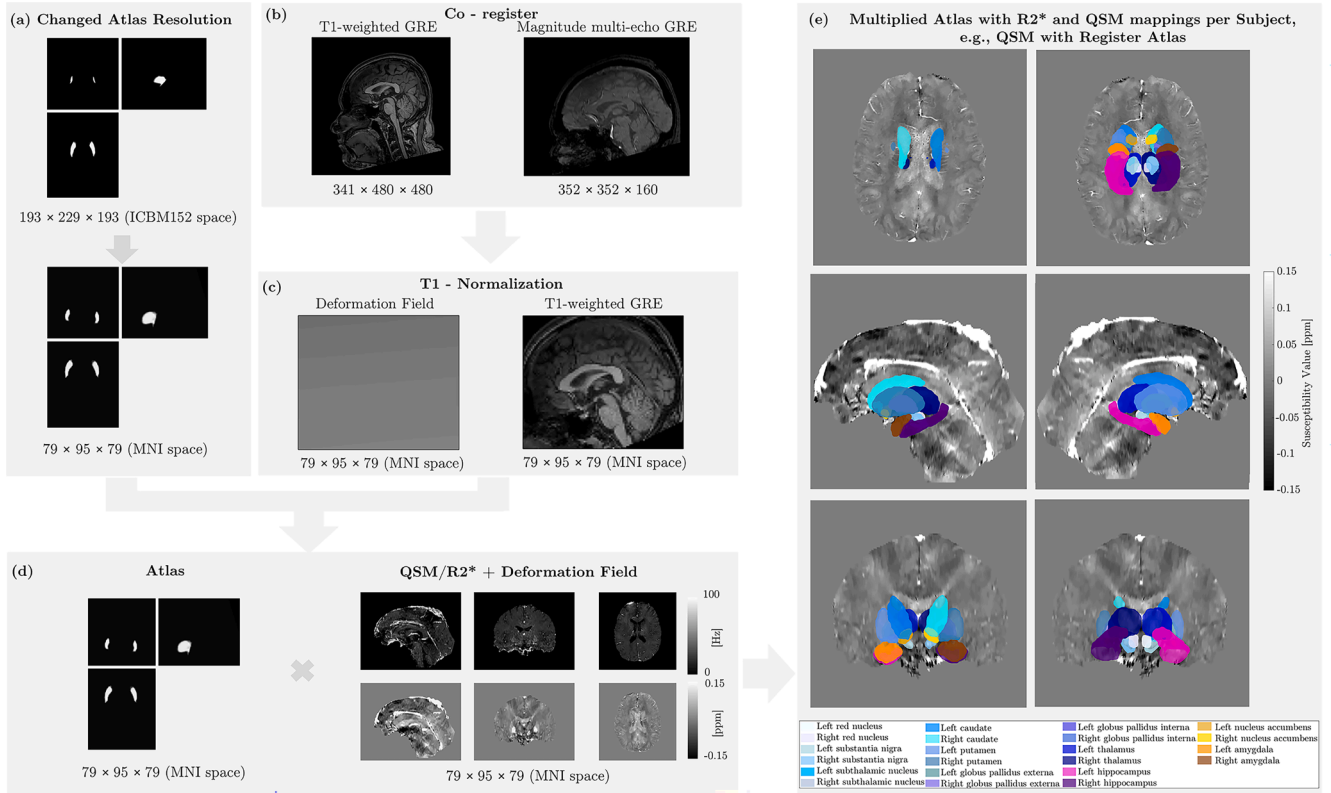


Figure 1. Schematic description of the registration process. (a) First, we changed the resolution of the atlas PD25 version 2019 (ICBM152 space) to the MNI space. (b) Second, we co-register the T1-weighted GRE and the magnitude multi-echo GRE. (c) Third, T1 was normalized, and the deformation field was obtained. (d) We applied the deformation field to the QSM and R2* maps and normalized them in the MNI space (e). Finally, we multiplied the atlas with QSM and R2* maps in each subject's space to obtain twenty-two mean magnetic susceptibility values per map and the same number of mean R2* values.

parameters are detailed in Table 1.

Also, we realized exploratory experiments using the same strategy mentioned but only using QSM or R2* for each classification task. Supplementary Table 1 showed the best hyperparameters found for each model analyzed.

2.5.4. Model evaluation

To ensure unbiased performance estimates and generalizability, we used stratified 10-fold cross-validation. This method splits the dataset into 10 subsets, training on 9 and validating on 1 in each iteration while accounting for class imbalance. Model performance was evaluated using accuracy, precision, recall, and F1-score on the test set.

2.5.5. Model interpretability

We calculated SHAP values to interpret the model predictions and understand the contribution of each feature [26]. TreeSHAP, a fast and precise method for estimating SHAP values in tree models, was used to explain the RF model. This approach combines local explanations of individual predictions to provide a global understanding of the model's structure, highlighting the importance and impact of each feature in classification tasks.

Table 1

Best hyperparameters found for the RF model for each combination of classes.

RF parameters	HV vs. FEP	RS vs. TRS
Number of Estimators	76	105
Maximum Depth	3	2
Maximum Leaf Nodes	3	2
Minimum Samples Split	19	12
Maximum Number of Features	Log2	Log2

2.6. QSM and R2* parameters analysis: hierarchical clustering

This study employed hierarchical clustering with Pearson correlation to analyze the structure of QSM and R2* features, comparing HV vs. FEP and RS vs. TRS patients. Hierarchical clustering, a data-driven method, groups similar features to facilitate dimensionality reduction, validate feature selection, and enhance model interpretability. Pairwise Pearson correlation coefficients were calculated to identify linear relationships between features, and the resulting correlation matrix served as the foundation for clustering analysis. The hierarchical clustering algorithm grouped features based on their correlations.

2.7. Statistical analysis

The statistical analyses were performed using R 4.1.1 (R Foundation for Statistical Computing, Vienna, Austria) [44]. We used the Shapiro-Wilk test to verify the normality of our demographic and clinical data. Considering the result of the previous tests, inter-group comparisons were analyzed using the student *T*-test or Mann-Whitney *U* test for normal or non-normal distributed data, respectively. All data were presented as mean $\pm$ standard deviation or median (range). The chi-square (χ^2) test was applied for categorical variables to assess the association between groups. This test is appropriate for examining the relationships between categorical data and helps determine if any observed differences between groups are statistically significant. A *p*-value < 0.05 was considered statistically significant.

3. Results

3.1. Demographics

Demographic and clinical data are described in Table 2. No significant difference was observed in gender composition between HV and FEP patients ($p = 0.097$). Sixty-eight percent of FEP patients did not have treatment resistance. We did not find any differences in age, sex, antipsychotic doses, and PANSS between patients with and without resistance to treatment.

Thirty-one percent of FEP patients presented non-response to at least two antipsychotic medications at appropriate doses (>600 mg chlorpromazine equivalents) for at least six weeks. They were at least moderately unwell (PANSS total > 58 , [33]). They did not have any prior history of exposure to clozapine.

3.2. QSM and R2* selection with SFS

Fig. 2 illustrates the importance of the feature derived from a linear model for distinguishing between groups: (a) HV vs. FEP patients and (b) RS vs. TRS. The importance values, represented on the y-axis, reflect the contribution of each feature to the model's decision-making process. Fig. 2.a, the most significant features for differentiating HV and FEP groups are primarily R2* values from specific brain regions, with the right nucleus accumbens R2* and left amygdala R2* showing the highest importance. Other notable areas include the left thalamus, left putamen, and right hippocampus, all contributing substantially to the model's performance. Interestingly, QSM-based features rank lower in importance, indicating that R2* measures play a dominant role in distinguishing these groups. Fig. 2.b, for the RS vs. TRS comparison, the feature importance distribution is more attenuated, with smaller overall contributions from individual features. The most important features are the right hippocampus R2*, left globus pallidus interna R2*, and right thalamus R2*. Compared to (a), QSM features again exhibit lower importance, suggesting that R2* features are more relevant for classification in both comparisons.

Fig. 3 presents the accuracy performance of the RF classifier based on the most relevant brain regions identified through SFS for both comparisons: (i) HV vs. FEP and (ii) RS vs. TRS. SFS selects the best single feature and then adds one feature at a time, maximizing the separability in combination with the selected features. Fig. 3.a. shows the accuracy performance obtained from the RF concerning the most relevant brain regions selected by SFS. Four variables (right nucleus accumbens R2*, left amygdala R2*, left nucleus accumbens R2*, and right thalamus QSM, Fig. 3.b) presented the maximum performance for the HV vs. FEP patient case, allowing us to obtain an accuracy of 76.48 ± 10.73 % (precision: 70.78 ± 9.39 %, recall: 67.35 ± 9.15 %, f1-score: 66.36 ± 9.68 %). For the case of RS vs. TRS: four variables were selected (right hippocampus R2*, left caudate R2*, left putamen R2*, and left amygdala QSM, Fig. 3.b), obtaining an accuracy of 76.43 ± 12.57 % (precision: 63.46 ± 25.59 %, recall: 64.83 ± 10.42 %, f1-score: 61.97 ± 10.17 %).

RF decision trees are shown in Supplementary Figure 1 for each

classification problem. Each node contains the feature ID and threshold used for splitting. The position of some features, e.g., the relative distance from the root, in the RF, reflected the strength of the association between QSM and R2* parameters and the FEP. The decision boundaries for the RF classifier are presented in Supplementary Figures 2 and 3. These visualizations demonstrate how the RF classifier separates the classes based on the selected features, with smoother and more complex decision surfaces due to the ensemble approach.

Supplementary Figure 4 shows the results for HV vs. FEP patients: 66.15 ± 14.11 % with four QSM features (right nucleus accumbens, left amygdala, left nucleus accumbens, Fig. 4.i) and 72.80 ± 9.37 % with five R2* features (left caudate, right caudate, left globus pallidus externa, right subthalamic nucleus, and left globus pallidus interna, Fig. 4.ii). Supplementary Figure 5 shows the results for RS vs. TRS patients: 75.36 ± 11.79 % with three QSM features (right caudate, right nucleus accumbens, and right amygdala, Fig. 5.i) and 75.89 ± 12.01 % with four R2* feature (left subthalamic nucleus, right nucleus accumbens, right amygdala, and left caudate, Fig. 5.ii).

3.3. QSM and R2* predictors of the model using SHAP analysis

After removing variables using the feature selection, treeSHAP was performed to find the most important predictors of the model, Fig. 4. In SHAP summary plots (Fig. 4.a), the impact of a feature on classes is stacked to create the feature importance graph, which shows what the RF learned from the features. Higher SHAP values indicated a more significant contribution to the model prediction. For HV vs. FEP, the most relevant features were right nucleus accumbens R2*, left amygdala R2*, left nucleus accumbens R2*, and right thalamus QSM. Whereas for RS vs. TRS patients, the most relevant features were left amygdala QSM, right hippocampus R2*, left caudate R2*, and left putamen R2*. For both classification problems, the HV vs. FEP and RS vs. TRS use the same features equally. That is why the confusion between classes is relatively high. One needs to generate a single feature dedicated to these classes to separate them better.

Fig. 4.b. shows the local explanation summary, combining feature importance with feature effects for our classification problems. Each point on the summary plot is a SHAP-value (x-axis) versus a QSM or R2* feature (y-axis). The least important feature has low SHAP values: the right thalamus QSM for HV vs. FEP, and the left putamen R2* for RS vs. TRS. Overlapping points are jittered in the y-axis direction, so we understand the distribution of the SHAP-values per QSM and R2* features. The features are ordered according to their importance in both classification problems. For example, in the HV vs FEP comparison, the right nucleus accumbens R2* emerges as the most significant variable. Lower values of this variable increase the likelihood of being classified as FEP. On the other hand, higher left amygdala R2* values promote the HV categorization and reduce the possibility of being categorized as FEP.

As an illustration of the possibilities of local explanations using treeSHAP, individual disease predictions are presented in Fig. 5. For our first study, HV vs. FEP patients (Fig. 5.a), two positive values contributed to the FEP class (left amygdala R2* and left nucleus accumbens

Table 2

Demographic and clinical data for HV and FEP patients. quantitative data are expressed as the median (range). PANSS: positive and negative syndrome scale. * Indicates statistically significant differences ($p < 0.05$). ** Antipsychotic doses expressed in chlorpromazine equivalents [46].

	HV			FEP Group		
	HV	FEP	p-value	RS	TRS	p-value
N	61	76		52	24	
Age [year]	23 (15,30)	20 (15,35)	$<0.001^*$	20 (16,35)	19 (15,25)	0.233
Sex [female: male]	40:21	60:16	0.097	39:13	21:3	0.332
Antipsychotic Doses** [mg]		450(25,1200)		431.3(30,975)	510(25,1200)	0.198
PANSS						
Positive		15 (7,43)		14(7, 43)	16 (10,21)	0.739
Negative		19.5 (7,41)		18(7,41)	25 (9,37)	0.015
General Psychopathology		30 (16,47)		29.5(16,47)	33 (20,41)	0.253
Total		66(39,100)		63.5(39,100)	72(50,97)	0.098

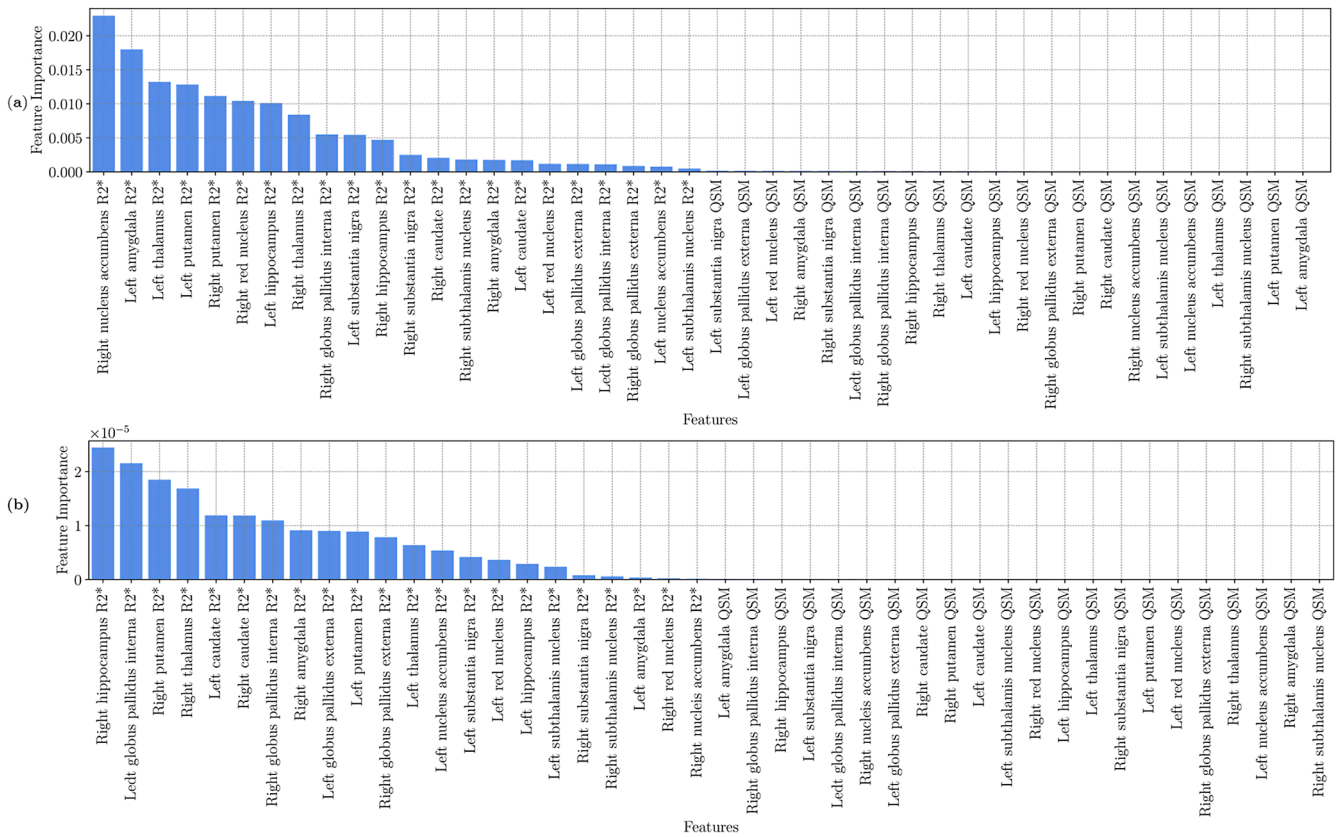


Figure 2. Feature importance using a linear model. For (a) HV vs. FEP patients and (b) RS vs. TRS.

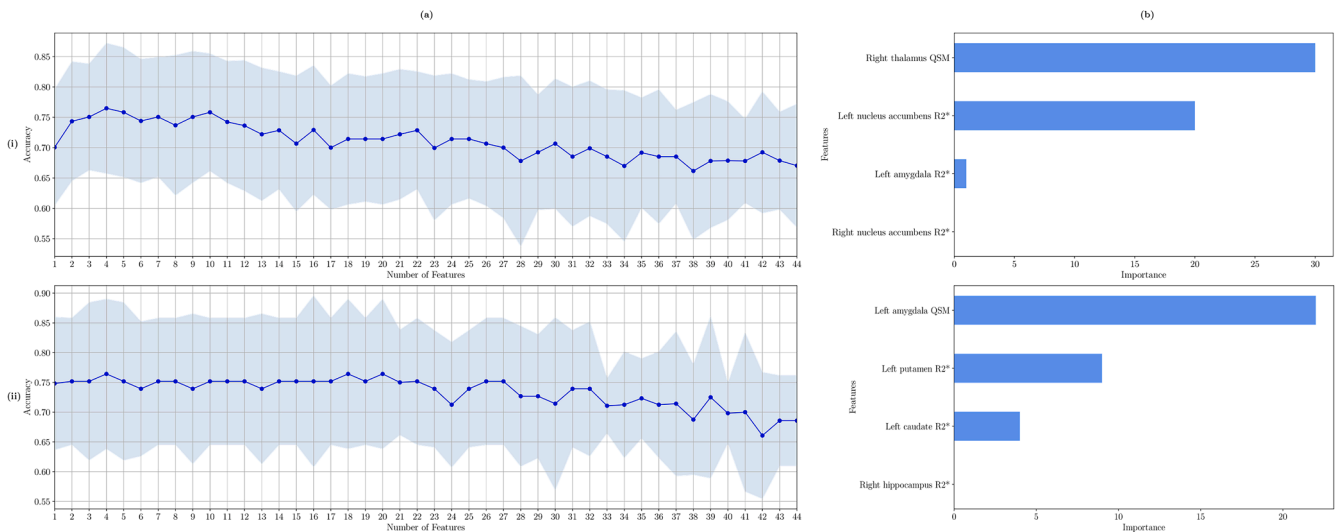


Figure 3. (a) Accuracy performance obtained from the RF concerning the most relevant brain regions selected by SFS. The maximum global accuracy of the RF for: (i) HV vs. FEP patients was 76.48 ± 10.73 % with (b) the four-best feature combination: right nucleus accumbens R2*, left amygdala R2*, left nucleus accumbens R2*, and right thalamus QSM and (i) RS vs. TRS was 76.43 ± 12.57 %, with (b) the four-best feature combination: right hippocampus R2*, left caudate R2*, left putamen R2*, and left amygdala QSM.

R2*), and two negative values contributed to the HV class (right nucleus accumbens R2* and right thalamus QSM). Regarding RS vs. TRS patients (Fig. 5.b), three negative values contributed to the RS class (left amygdala QSM, left caudate R2*, and left putamen R2*). In contrast, one positive value contributed to the TRS class (right hippocampus R2*). Fig. 6 presents the brain areas found as disease predictors by treeSHAP per map and classification problem.

3.4. QSM and R2* parameters correlation

Finally, an analysis of the hierarchical clustering, Fig. 7, shows that QSM and R2* features were divided into two clusters (red and green) for both classes' analysis (HV vs. FEP and RS vs. TRS). Both clusters were combined at a high dendrogram distance and thus, can be treated as individual analysis groups. For HV and FEP patients, we found cluster 1 (red): R2* values of left and right red nucleus, caudate, putamen, globus

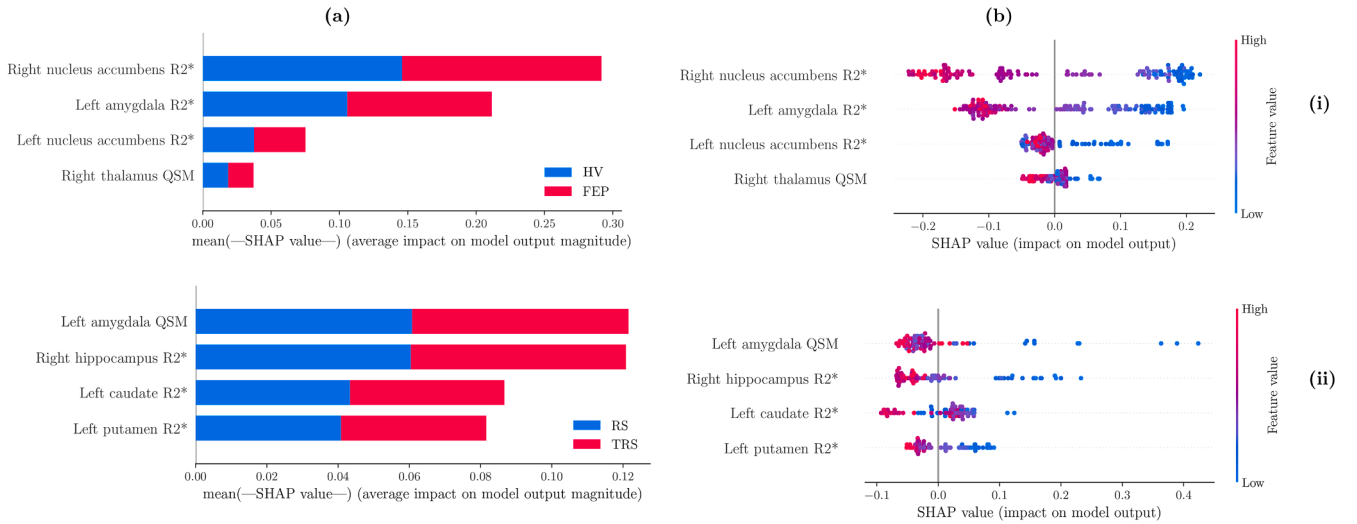


Figure 4. Feature the importance of our models based on SHAP values. (a) The mean absolute SHAP values are depicted to illustrate the importance of global features. (b) The local explanation summary shows the direction of the relationship between a variable and class study. (i) Positive SHAP values push the model to predict FEP patients, and negative values push the model to predict the HV. (ii) Positive SHAP values predict RS patients, and negative values predict TRS patients.

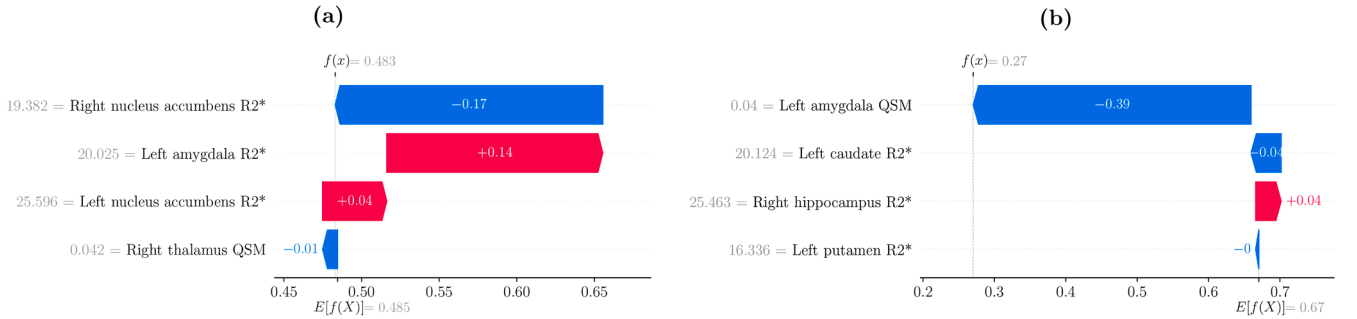


Figure 5. Waterfall plot explains local disease prediction and shows the contribution of each variable to the forecast. The forecast of HV vs. FEP patients, from the perspective of HV patients, is shown in (a). The forecast of RS vs. TRS patients, from the perspective of RS patients, is shown in (b). The value of each variable precedes the variable name. Below the x-axis, the baseline value ($E[f(X)]$) is displayed, indicative of the expected value of the model evaluated on the background dataset. The SHAP values of each variable are summed to match the model output with all variables included. Positive SHAP values push the model to predict HV or RS, and negative values predict FEP or TRS patients depending on the classification problem.

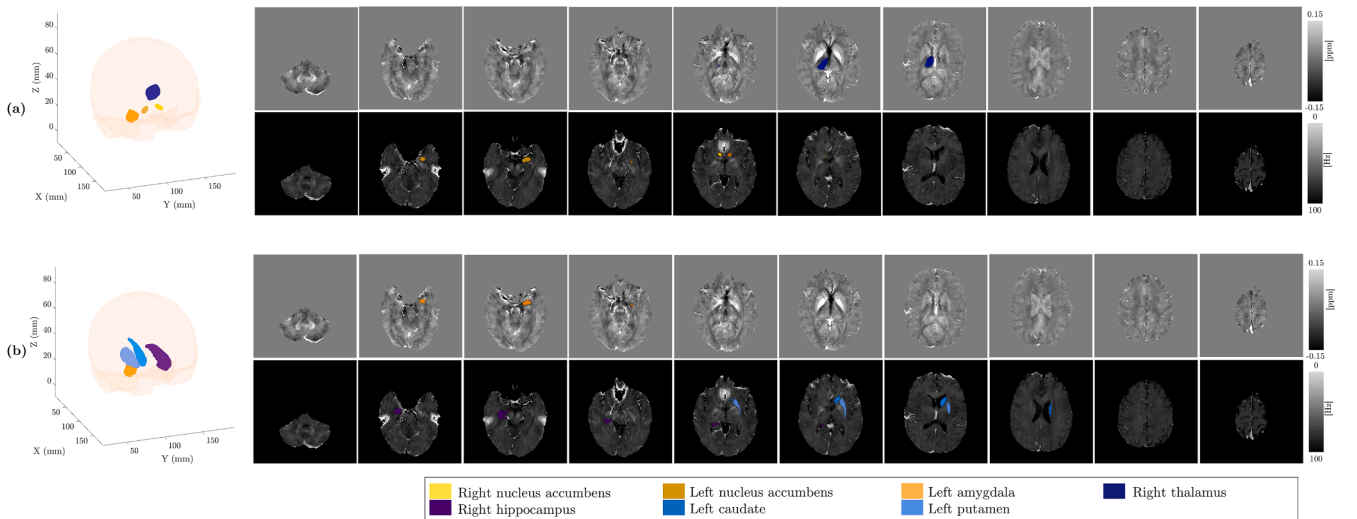


Figure 6. Discriminating brain regions according to treeSHAP per map (QSM and R2*). (a) HV vs. FEP patients and (b) RS vs. TRS patients. 3D rendering and the axial plane of the selected regions' QSM and R2* maps.

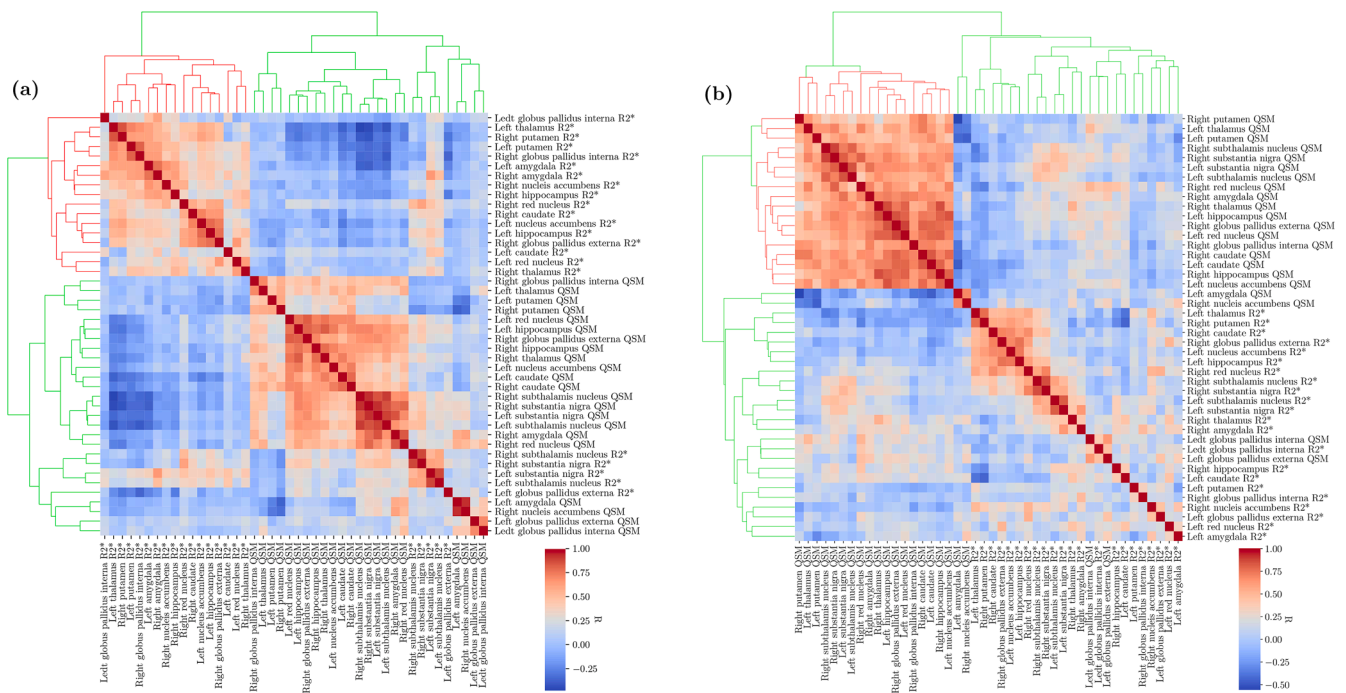


Figure 7. Dendrogram and hierarchical clustering results based on average linkage method for all magnetic susceptibility parameters of (a) HV and FEP patients and (b) RS and TRS patients.

pallidus interna, thalamus, hippocampus, nucleus accumbens, amygdala, and right globus pallidus externa; and cluster 2 (green): R2* mapping of right and left substantia nigra and subthalamic nucleus, and left globus pallidus externa; and QSM mapping of left and right red nucleus, substantia nigra, subthalamic nucleus, caudate, putamen, globus pallidus externa and interna, thalamus, hippocampus, nucleus accumbens, and amygdala.

For RS vs. TRS patients, cluster 1 (red): QSM mapping of right and left thalamus, putamen, caudate, red nucleus, hippocampus, substantia nigra, subthalamic nucleus, left nucleus accumbens, right thalamus, right globus pallidus externa, and right amygdala; and Cluster 2 (green): QSM mapping of left globus pallidus externa and interna, right nucleus accumbens, and left amygdala; and R2* mapping of left and right red nucleus, substantia nigra, subthalamic nucleus, caudate, putamen, globus pallidus externa, and interna, thalamus, hippocampus, nucleus accumbens, and amygdala.

4. Discussion

Iron plays a crucial role in the normal functioning and development of the brain. However, excess or deficit iron deposition has been associated with brain pathologies [45]. Iron molecules are essential for several metabolic pathways, particularly in synthesizing dopamine [46]. QSM and $R2^*$ values are important for evaluating iron accumulation, which has been studied in deep gray nuclei in patients with psychosis [11]. Recent studies suggest a relationship between iron deposition, psychosis, and blood-brain barrier (BBB) dysfunction, which may play a role in the pathophysiology of schizophrenia [47]. Our findings align with these observations.

This study evaluated potential differences in dopamine-related deep brain nuclei between HV and FE patients. We used an ML algorithm (with RF models) to identify the most important areas that allowed us to discriminate between those groups.

We proposed two RF models. With the first one we identified the most relevant nuclei whose QSM or R2* values allow us to discriminate between patients with FEP and healthy subjects. With the second one, we identified the most relevant nuclei whose QSM or R2* values allow

us to discriminate between patients who respond or do not respond to antipsychotics. Given the large number of potential descriptors for classifying FEP patients, we employed feature selection algorithms to reduce the number of input variables, focusing on the most significant regions for accurate classification. More importantly, we aimed to open the black box of the model and explain the prediction results of the model using the SHAP methodology. This tree explainer uses SHAP values and allowed us global and local interpretability, leading to understanding the model behavior [11,21].

With the four-best features combination, our approach accurately classified HV vs. FEP patients and RS vs. TRS patients with 76.48 ± 10.73 % (based on $R2^*$ values in the nucleus accumbens, amygdala, and QSM in the thalamus) and 76.43 ± 12.57 % accuracy (based on $R2^*$ values in the hippocampus, caudate, putamen, and QSM in the amygdala), respectively. However, the accuracy performance decreased when we considered the ML strategy using only a type of feature. For HV vs. FEP patients, we obtained 66.15 ± 14.11 % with four QSM features and 72.80 ± 9.37 % with five $R2^*$ features, and for RS vs. TRS patients, we got 75.36 ± 11.79 % with three QSM features and 75.89 ± 12.01 % with four $R2^*$ features. Both provide different contrasts and allow us to identify changes in paramagnetic and diamagnetic concentrations differently. Therefore, both classification problems performed better when including both QSM and $R2^*$ values than independently. Although we used 10-fold stratified cross-validation, we did not repeat experiments with multiple random seeds. This may affect stability, and future work should incorporate repeated random sub-sampling validation.

We developed our ML with the understanding that RF is particularly adept at capturing complex relationships within the data. To enhance model performance, we conducted a hyperparameter grid search using stratified cross-validation. However, the high standard deviation suggests that the model is sensitive to data imbalance, which can undermine its stability. While stratified sampling can mitigate some issues, it may not fully resolve the variability in performance across different folds, as certain folds may still have fewer instances of the minority class, negatively impacting the model's generalization ability. Additionally, the inherent complexity of the classification task can introduce further variability; even small changes in the data can lead to significant

differences in model performance.

To address these challenges, future modeling efforts should adopt a multi-faceted approach to enhance robustness and improve results. We can start by exploring advanced oversampling techniques such as SMOTE (Synthetic Minority Over-sampling Technique) or ADASYN to bolster the representation of the minority class. Additionally, implementing robust RF configurations, feature selection, and hyper-parameter tuning would optimize model performance.

Additionally, treeSHAP that uses SHAP-values formulation guarantees three important properties: local accuracy, missingness, and consistency [23]. Therefore, its results can be interpreted as the feature's importance and reflect its influence on the prediction. Considering the relative importance of independent variables, the treeSHAP found the four best predictive variables and their influence on each class. Considering the relative significance of independent variables, treeSHAP identified the four most effective predictive variables for HV vs. FEP classes (R2*: right nucleus accumbens, left amygdala, left nucleus accumbens, and QSM: right thalamus). Notably, we observed higher R2* values and a positive contribution of SHAP values in the left amygdala and nucleus accumbens for FEP patients, while HV subjects exhibited lower R2* and QSM values and a negative contribution for the right nucleus accumbens and thalamus, respectively. These findings not only enhance our understanding of these brain regions but also underscore the importance of our research in the context of schizophrenia. The amygdala, a well-known area that has been extensively studied, plays a pivotal role in schizophrenia. Its involvement in emotion is profound, as it is the area that, when stimulated, can evoke complex hallucinations and emotions [48]. This deep understanding of the amygdala's role adds a layer of complexity to our research and the field of neuroscience. The nucleus accumbens, a key component of the mesolimbic dopamine system, is responsible for the brain's reward pathway. Its role in reinforcement learning, everyday behaviors, and its potential contribution to altered salience processing in schizophrenia is an area of great interest and curiosity [49]. This potential impact of our research on understanding altered salience processing in schizophrenia is significant and should not be underestimated. The thalamus is well known to be involved in high-order information processing. This area is the gateway to cortical processing for incoming sensory information. Basal ganglia modulate the activity of thalamic projections, cortico-striato-pallido-thalamo-cortical loop. Increased stimulation of sensory cortical regions, along with dysfunction in cortical association areas, particularly the dorsolateral prefrontal cortex [50].

Furthermore, the four most influential features (QSM: left amygdala and R2*: right hippocampus, left caudate, and left putamen) in SHAP contribution for RS vs. TRS classes have been previously investigated in cohorts with treatment-resistant schizophrenia (TRS). For example, Mier et al. observed distinct activation and connectivity patterns in the amygdala, indicating a potential neural mechanism in schizophrenia patients [51]. These patients exhibited reduced functional connectivity in the hippocampus, fusiform gyrus, and middle temporal gyrus, suggesting an impairment in the integration of memory-informed processing streams by the right temporoparietal junction during mental state inferences. The coupling of the hippocampus played a role in the participants' adaptive social behavior, implying that a neural pathway crucial for updating social knowledge and making forward predictions in social interactions is altered in schizophrenia [51]. Additionally, abnormalities in hemispheric specialization were noted in the caudate nucleus and cortical regions connected to it in schizophrenia [52]. Huang et al. also discovered a correlation between decreased left putamen and delusions in schizophrenia patients [53].

The hierarchical clustering provided a method for a reliable correlation between QSM and R2* features. The resulting dendrogram represents the relationship between QSM and R2* subcortical areas by classifying them into two clusters according to their similarities: one cluster mostly conformed by QMS data and the other by R2*. In that context, O'Brien showed that multicollinearity is not a sufficient reason

to eliminate variables from a model. Instead, a more important criterion for dropping a variable is model influence [54]. Furthermore, Gregorich et al. showed that two highly correlated independent variables would lead to high variance in the predictions, even if both variables are relevant for prediction. Therefore, omitting one of the pairs in small samples may be beneficial to decrease that variance, even if this incurs some new bias in the predictions [55]. Even though we studied the features correlations and clustering, this information was not used to intervene in the ML model since the latter selects the features automatically. Instead, we used this information to explain the relation and locations across the different cluster features chosen by our ML approach. This important information on relative performance can be used to aid researchers in selecting appropriate areas to compute susceptibility and R2* maps for their studies.

R2* and QSM maps are to some extent related to each other. Indeed, $R2^* = R2 + R2'$, where $R2'$ is the transversal relaxation decay proportional to the absolute value of susceptibility. Therefore, each region should be highly correlated in both maps, which was proved in the hierarchical clustering analysis. Nevertheless, they could have different predictive powers depending on the sensitivity to noise and the type of chemical change involved.

The technical novelty of the present study lies in its integration of advanced neuroimaging techniques, particularly QSM and R2* values, with methods for predicting FEP and response to antipsychotic treatments. A key innovation of this research is introducing an RF model, which incorporates these neuroimaging biomarkers to predict treatment outcomes in FEP patients. While ML has been applied to neuroimaging data in other contexts, our approach presents a novel application by leveraging RF models to discern patterns in QSM and R2* data specifically for early psychosis prediction and treatment response.

A particularly innovative aspect of this study is the use of SHAP to enhance the interpretability of the RF model. By explaining the importance of individual features in the model's predictions, SHAP addresses a common challenge in ML-based models: the difficulty of understanding how predictions are made. This addition is crucial, as it allows clinicians and researchers to gain insight into the relationships between neuroimaging biomarkers and clinical outcomes, offering a more transparent and trustworthy predictive framework for psychosis.

Moreover, this study also employs FS techniques to reduce the complexity of the dataset and improve model performance. By identifying the most relevant biomarkers, FS ensures that the RF model remains efficient without sacrificing predictive power. This is particularly important given the high dimensionality of neuroimaging data, where thousands of variables could otherwise overwhelm traditional models. This streamlined approach reduces the computational burden and improves model generalization, ensuring the predictive results are robust and applicable to new datasets.

4.1. Limitations

Our study has several limitations that should be considered when interpreting our findings. First, the small sample size of our cohorts limits the generality of our findings [56]. Although we used 10-fold stratified cross-validation, we did not repeat the experiments with multiple random seeds. This may affect stability, so future work should incorporate repeated validation with random sub-samples to ensure more robust estimates.

Second, although we recorded the duration of antipsychotic exposure before MRI (mean $\pm$ SD: 36 $\pm$ 46.5 days) and considered it a potential confounder, we did not include this variable as a covariate in our models. The wide variability in treatment duration and the small sample size limited our ability to formally adjust for this factor. Future studies with larger cohorts should explicitly model treatment duration, either as a covariate or through stratified analyses, to clarify its influence on QSM and R2* measures.

Third, the model was developed and tested in a single-site cohort

from Chile without external validation. While this ensured methodological consistency, it constrains external validity. Multicenter studies with larger and more heterogeneous cohorts are required to confirm the reproducibility and clinical applicability of our results in diverse populations.

Fourth, the identification of TRS relied on a combination of clinical reassessments and data from the national pharmacovigilance program (using clozapine prescription as a proxy for TRS). We may have missed some TRS patients because they were not clinically assessed, and had not yet started clozapine treatment.

Fifth, due to the small sample size, we did not explore the use of advanced deep learning algorithms. Instead, we used classical ML algorithms, which nevertheless achieved high classification accuracy in both studies. Deep networks require large training datasets to achieve high performance avoid overfitting. In future studies, we will include more participants to apply advanced deep learning methods, apply regularization techniques to reduce model complexity, and compare them with classical ML algorithms. Longitudinal studies are also needed to investigate the prognostic relevance of our model.

Sixth, hyperparameter optimization and model evaluation were performed on the same dataset using cross-validation, which could introduce bias. A more rigorous approach would include nested cross-validation or a separate validation dataset to better estimate generalization performance.

Finally, QSM and $R2^*$ maps were analyzed without explicitly separating paramagnetic and diamagnetic contributions. In future studies, we will explore algorithms such as X-separation (chi-separation) or DECOMPOSE [57] to separate the paramagnetic and diamagnetic components in each voxel and test whether this refinement improves classification performance.

5. Conclusions

We identified four brain nuclei whose QSM and $R2^*$ features are critical for classification, utilizing a feature selection algorithm (SFS) and tree explainer (treeSHAP). For distinguishing HV from FEP patients, the relevant features included the $R2^*$ values from the right nucleus accumbens, left amygdala, and left nucleus accumbens, along with the QSM value from the right thalamus, achieving an accuracy of 76.48 ± 10.73 %. The significant features for the RS versus TRS classification included the $R2^*$ values from the right hippocampus, left caudate, left putamen, and the QSM value from the left amygdala, with an accuracy of 76.43 ± 12.57 %.

Moreover, the hierarchical clustering method revealed a reliable correlation between QSM and $R2^*$ features, identifying two distinct clusters relevant to our classification challenges.

The features we identified could be instrumental in future studies for early detection of treatment-resistant FEP patients. This could enable the timely initiation of appropriate treatments, such as clozapine, with the potential to significantly enhance clinical outcomes and reduce the functional disability and social burden associated with prolonged psychosis during a critical period.

Funding

This work has been funded by the National Agency for Research and Development (ANID), project ICN2021_004 of the Millennium Science Initiative Program. CT was also funded by Fondecyt 1231,535. The author RCC would like to thank the financial support provided by ANID through the project Fondecyt de Iniciación en Investigación 2025 N° 11250867, by the Competition for Research Regular Projects, year 2023, code LPR23-17, Universidad Tecnológica Metropolitana, by the Internal Research Assistant Funding Competition UTEM, year 2025, code AI25-07, Universidad Tecnológica Metropolitana, and by the Competition for Scientific and Technological Equipment Projects, year 2024, code LE24-03, Universidad Tecnológica Metropolitana. Project supported by

the Translational Biomedical Data, Signals and Imaging Laboratory (LaTSIB).

Data availability

The machine learning pipelines and visualization scripts developed for this study are available in a private GitHub repository for reviewer access at: <https://github.com/pamelaFranco/Interpretable-FEP-QSM>. The datasets generated and/or analyzed during the current study are not publicly available due to data privacy restrictions imposed by the Ethics Committee of the Pontificia Universidad Católica de Chile (Ref: 15–297) and the Servicio de Salud Metropolitano Norte. Data may be made available from the corresponding author (Dr. Nicolás Crossley) upon reasonable request and subject to institutional approval.

Ethical statement

The study was approved by the Ethics Committee of the *Pontificia Universidad Católica de Chile* (Ref: 15–297) and *Servicio de Salud Metropolitano Norte*.

CRediT authorship contribution statement

Pamela Franco: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Cristian Montalba:** Writing – review & editing, Formal analysis. **Raúl Caulier-Cisterna:** Writing – review & editing, Formal analysis. **Carlos Milovic:** Writing – review & editing, Software. **Alfonso González:** Writing – original draft, Funding acquisition. **Juan Pablo Ramirez-Mahaluf:** Writing – original draft, Funding acquisition. **Juan Undurraga:** Writing – original draft, Funding acquisition. **Rodrigo Salas:** Writing – original draft, Supervision. **Nicolás Crossley:** Writing – original draft, Supervision, Funding acquisition. **Cristian Tejos:** Writing – review & editing, Supervision. **Sergio Uribe:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no competing interests.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.cmpb.2025.109067](https://doi.org/10.1016/j.cmpb.2025.109067).

References

- [1] P.B. Jose, Adult mental health disorders and their age at onset, *Br. J. Psychiatry* 54 (2013) s5–10, <https://doi.org/10.1192/bjp.bp.112.119164>. Suppl.
- [2] GBD 2017. (2018). Disease and injury incidence and prevalence collaborators. 392 (10159):1789–1858. [doi: 10.1016/S0140-6736\(18\)32279-7](https://doi.org/10.1016/S0140-6736(18)32279-7).
- [3] F. Iasevoli, S. Giordano, R. Balletta, G. Latte, M.V. Formato, E. Prinziavalli, D. De Berardis, C. Tomasetti, A. de Bartolomeis, Treatment resistant schizophrenia is associated with the worst community functioning among severely-ill highly-disabling psychiatric conditions and is the most relevant predictor of poorer achievements in functional milestones, *Prog. Neuropsychopharmacol. Biol. Psychiatry* 65 (2016) 34–48, <https://doi.org/10.1016/j.pnpbp.2015.08.010>.
- [4] S. Chakrabarti, Clozapine resistant schizophrenia: newer avenues of management, *World J. Psychiatry* 11 (8) (2021) 429–448, <https://doi.org/10.5498/wjp.v11.i8.429>. Aug 19 PMID: 34513606; PMCID: PMC8394694.
- [5] O. Agid, G. Remington, S. Kapur, T. Arenovich, R.B. Zipursky, Early use of clozapine for poorly responding first-episode psychosis, *J. Clin. Psychopharmacol.* 27 (4) (2007 Aug) 369–373, <https://doi.org/10.1097/jcp.0b013e3180d0a6d4>. PMID: 17632221.
- [6] O.D. Howes, J. Kambeitz, E. Kim, D. Stahl, M. Slifstein, A. Abi-Dargham, S. Kapur, The nature of dopamine dysfunction in schizophrenia and what this means for treatment, *Arch. Gen. Psychiatry* 69 (8) (2012) 776–786, <https://doi.org/10.1001/archgenpsychiatry.2012.169>.
- [7] C.J. Bachmann, L. Aagaard, M. Bernando, L. Brandt, M. Cartabia, A. Clavenna, A. Coma Fusté, K. Furu, K. Garuoliene, F. Hoffmann, S. Hollingworth, K. F. Huybrechts, L.J. Kalverdijk, K. Kawakami, H. Kieler, T. Kinoshita, S.C. López, J. E. Machado-Alba, M.E. Machado-Duque, M. Mahesri, P.S. Nishtala, D. Piovani,

- J. Reutfors, L.K. Saastamoinen, I. Sato, C.C.M. Schuiling-Veninga, Y.C. Shyu, D. Siskind, S. Skurtveit, H. Verdoux, L.J. Wang, C. Zara Yahni, H. Zoëga, D. Taylor, International trends in clozapine use: a study in 17 countries, *Acta Psychiatr. Scand.* 136 (1) (2017) 37–51, <https://doi.org/10.1111/acps.12742>.
- [8] Nikolai Albert, Melissa Authen Weibell, The outcome of early intervention in first episode psychosis, *Int. Rev. Psychiatry* 31 (5–6) (2019) 413–424, <https://doi.org/10.1080/09540261.2019.1643703>.
- [9] S.M. Stahl, Beyond the dopamine hypothesis of schizophrenia to three neural networks of psychosis: dopamine, serotonin, and glutamate, *CNS. Spectr.* 23 (3) (2018) 187–191, <https://doi.org/10.1017/S1092852918001013>.
- [10] A. Blazer, K.N.R. Chengappa, W. Foran, A.C. Parr, C.E. Kahn, B. Luna, D.K. Sarpal, Changes in corticostriatal connectivity and striatal tissue iron associated with efficacy of clozapine for treatment-resistant schizophrenia, *Psychopharmacology* 239 (8) (2022 Aug) 2503–2514, <https://doi.org/10.1007/s00213-022-06138-0>. Epub 2022 Apr 18. PMID: 35435461; PMCID: PMC9013738.
- [11] M. Xu, Y. Guo, J. Cheng, K. Xue, M. Yang, X. Song, Y. Feng, J. Cheng, Brain iron assessment in patients with first-episode schizophrenia using quantitative susceptibility mapping, *Neuroimage Clin.* 31 (2021) 102736, <https://doi.org/10.1016/j.nicl.2021.102736>.
- [12] L. De Peri, A. Crescini, G. Deste, P. Fusar-Poli, E. Sacchetti, A. Vita, Brain structural abnormalities at the onset of schizophrenia and bipolar disorder: a meta-analysis of controlled magnetic resonance imaging studies, *Curr. Pharm.* 18 (2012) 486–494, <https://doi.org/10.2174/138161212799316253>.
- [13] Wheeler Al, A.N. Voineskos, A review of structural neuroimaging in schizophrenia: from connectivity to connectomics, *Front Hum. Neurosci.* 8 (2014) 653, <https://doi.org/10.3389/fnhum.2014.00653>.
- [14] Baker Jt, Dillon DG, L.M. Patrick, J.L. Roffman, R. O Brady Jr, D.A. Pizzagalli, D. Öngür, A.J. Holmes, Functional connectomics of affective and psychotic pathology, *Proc. Natl. Acad. Sci.* 116 (2019) 9050–9059, <https://doi.org/10.1073/pnas.1820780116>.
- [15] M.F. Casanova, S.O. Comparini, R.W. Kim, J.E. Kleinman, Staining intensity of brain iron in patients with schizophrenia: a postmortem study, *J. Neuropsychiatry Clin. Neurosci.* Winter 4 (1) (1992) 36–41, <https://doi.org/10.1176/jnp.4.1.36>. PMID: 1627959.
- [16] M.F. Casanova, I.N. Waldman, J.E. Kleinman, A postmortem quantitative study of iron in the globus pallidus of schizophrenic patients, *Biol. Psychiatry* 27 (2) (1990) 143–149, [https://doi.org/10.1016/0006-3223\(90\)90644-h](https://doi.org/10.1016/0006-3223(90)90644-h). Jan 15PMID: 2294978.
- [17] K.W. Lange, J. Kornhuber, P. Kruzik, W.D. Rausch, E. Gabriel, K. Jellinger, P. Riederer, Brain iron and schizophrenia, in: P. Riederer, M.B.H. Youdim (Eds.), *Iron in Central Nervous System Disorders. Key Topics in Brain Research*, Springer, Vienna, 1993, https://doi.org/10.1007/978-3-7091-9322-8_3.
- [18] Y. Wang, T. Liu, Quantitative susceptibility mapping (QSM): decoding MRI data for a tissue magnetic biomarker, *Magn. Reson. Med.* 73 (1) (2015) 82–101, <https://doi.org/10.1002/mrm.25358>.
- [19] P. Ravanfar, S.M. Loi, W.T. Syeda, T.E. Van Rhee, A.I. Bush, P. Desmond, V. L. Cropley, D.J.R. Lane, C.M. Opazo, B.A. Moffat, D. Velakoulis, C. Pantelis, Systematic Review: quantitative susceptibility mapping (QSM) of brain iron profile in neurodegenerative diseases, *Front. Neurosci.* Feb. 18 (15) (2021) 618435, <https://doi.org/10.3389/fnins.2021.618435>. PMID: 33679303; PMCID: PMC7930077.
- [20] C. Langkammer, F. Schweser, N. Krebs, A. Deistung, W. Goessler, E. Scheurer, K. Sommer, G. Reishofer, K. Yen, F. Fazekas, S. Ropele, J.R. Reichenbach, Quantitative susceptibility mapping (QSM) as a means to measure brain iron? A post mortem validation study, *Neuroimage* 62 (3) (2012 Sep) 1593–1599, <https://doi.org/10.1016/j.neuroimage.2012.05.049>. Epub 2012 May 24. PMID: 22634862; PMCID: PMC3413885.
- [21] M. García-Saborit, N. Muñoz, C. Milovic, A. Tepper, L.M. Allende, C. Mena, B. Iruretagoyena, J.P. Ramirez-Mahaluf, C. Diaz, R. Nachar, C.P. Castañeda, A. González, J. Undurraga, N. Crossley, C. Tejos, Quantitative susceptibility mapping MRI in deep brain nuclei in first episode psychosis, *Schizophr. Bull.* (2023), <https://doi.org/10.1093/schbul/sbad041>.
- [22] D.P. Kroese, *Data Science and Machine Learning: Mathematical and Statistical Methods*, 1st edition, Chapman and Hall/CRC, 2019.
- [23] S.M. Lundberg, G. Erion, H. Chen, A. DeGrave, J.M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansl, S.I. Lee, From local explanations to global understanding with explainable AI for trees, *Nat. Mach. Intell.* 2 (2020) 56–67, <https://doi.org/10.1038/s42256-019-0138-9>.
- [24] D. Bzdok, N. Altman, M. Krzywinski, Points of significance: statistics versus machine learning, *Nat. Methods* 15 (2018) 233–234, <https://doi.org/10.1038/nmeth.4642>.
- [25] D.B. Dwyer, C. Cabral, L. Kamberitz-Llankovic, R. Sanfelici, J. Kambeitz, V. Calhoun, P. Falkai, C. Pantelis, E. Meisenzahl, N. Koutsouleris, Brain subtyping enhances the neuroanatomical discrimination of schizophrenia, *Schizophr. Bull.* 44 (2018) 1060–1069, <https://doi.org/10.1093/schbul/sby008>.
- [26] S. Vieira, Q.Y. Gong, W.H.L. Pinaya, C. Scarpazza, S. Togni, B. Crespo—Facorro, D. Rodesillas-Gutierrez, V. Ortiz-Garcia, E. Setien-Suero, F.E. Scheepers, N.E. M. Van Haren, T.R. Marques, D.L. Murray, A. David, P. Dazzan, P. McGuire, A. Mechelli, Using machine learning and structural neuroimaging to detect first episode psychosis: reconsidering the evidence, *Schizophr. Bull.* 46 (1) (2020) 17–26, <https://doi.org/10.1093/schbul/sby189>.
- [27] D. Mery, *Computer Vision for X-Ray Testing: Imaging, Systems, Image Databases, and Algorithms*, Springer, 2015 first ed.
- [28] L. Breiman, Random forests, *Mach. Learn.* 45 (2001) 5–32, <https://doi.org/10.1023/A:1010933404324>.
- [29] D.V. Sheehan, K.H. Sheehan, R.D. Shytle, J. Janavs, Y. Bannan, J.E. Rogers, K. M. Milo, S.L. Stock, B. Wilkinson, Reliability and validity of the mini international neuropsychiatric interview for children and adolescents (MINI-KID), *J. Clin. Psychiatry* 71 (3) (2010) 313–326, <https://doi.org/10.4088/JCP.09m05305whi>.
- [30] S.R. Kay, A. Fiszbein, L.A. Opler, The positive and negative syndrome scale (PANSS) for schizophrenia, *Schizophr. Bull.* 13 (2) (1987) 261–276, <https://doi.org/10.1093/schbul/13.2.261>.
- [31] R.S. Kern, K.H. Nuechterlein, M.F. Green, L.E. Baade, W.S. Fenton, J.M. Gold, R.S. E. Keefe, R. Meshulam-Gately, J. Mintz, L.J. Seidman, E. Stover, S.R. Marder, The MATRICS consensus cognitive battery, part 2: co-norming and standardization, *Am. J. Psychiatry* 165 (2) (2008) 214–220, <https://doi.org/10.1176/appi.ajp.2007.07010043>.
- [32] K.H. Nuechterlein, M.F. Green, R.S. Kern, L.E. Baade, D.M. Barch, J.D. Cohen, S. Essock, W.S. Fenton, F.J. Frese 3rd, J.M. Gold, T. Goldberg, R.K. Heaton, R.S. E. Keefe, H. Kraemer, R. Meshulam-Gately, L.J. Seidman, E. Stover, D. R. Weinberger, A.S. Young, S. Zalcman, S.R. Marder, The MATRICS consensus cognitive battery, part 1: test selection, reliability, and validity, *Am. J. Psychiatry* 165 (2) (2008) 203–213, <https://doi.org/10.1176/appi.ajp.2007.07010042>.
- [33] S. Leucht, M. Samara, S. Heres, J.M. Davis, Dose equivalents for antipsychotic drugs: the DDD method, *Schizophr. Bull.* 1 (Suppl 1) (2016) S90–S94, <https://doi.org/10.1093/schbul/sbv167>, 42 Suppl.
- [34] O.D. Howes, R. McCutcheon, O. Adig, A. de Bartolomeis, N.J.M. van Beveren, M. L. Birnbaum, M.A.P. Bloomfield, R.A. Bressan, R.W. Buchanan, W.T. Carpenter, D. J. Castle, L. Citrome, Z.J. Daskalakis, M. Davidson, R.J. Drake, S. Dursun, B. H. Eddrup, H. Elkis, P. Falkai, W.W. Fleischacker, A. Gadelha, F. Gaughan, B. Y. Glenthøj, A. Graff-Guerrero, J.E.C. Hallak, W.G. Honer, J. Kennedy, B.J. Kinon, S.M. Lawrie, J. Lee, F.M. Leweke, J.H. MacCabe, C.B. McNabb, H. Meltzer, H. J. Möller, S. Nakajim, C. Pantelis, T.R. Marques, G. Remington, S.L. Rossell, B. R. Russell, C.O. Siu, T. Suzuki, I.E. Sommer, D. Taylor, N. Thomas, A. Üçok, D. Umbricht, J.T.R. Walters, J. Kane, C.U. Correll, Treatment-resistant schizophrenia: treatment response and resistance in psychosis (TRIP) working group consensus guidelines on diagnosis and terminology, *Am. J. Psychiatry* 174 (3) (2017) 216–229, <https://doi.org/10.1176/appi.ajp.2016.16050503>, 1.
- [35] B. Dymerska, K. Eckstein, B. Bachrata, B. Siow, S. Trattinnig, K. Shmueli, S. D. Robinson, Phase unwrapping with a rapid open-source minimum spanning tree algorithm (ROME), *Magn. Reson. Med.* 85 (4) (2021) 2294–2308, <https://doi.org/10.1002/mrm.28563>.
- [36] T. Liu, I. Khalidov, L. de Rochefort, P. Spincemaille, J. Liu, A.J. Tsiouris, Y. Wang, A novel background field removal method for MRI using projection onto dipole fields (PDF), *NMR Biomed.* 24 (2011) 1129–1136, <https://doi.org/10.1002/nbm.1670>.
- [37] C. Milovic, B. Bilgic, B. Zhao, C. Langkammer, C. Tejos, J. Acosta-Cabrero, Weak-harmonic regularization for quantitative susceptibility mapping, *Magn. Reson. Med.* 81 (2) (2019) 1399–1411, <https://doi.org/10.1002/mrm.27483>.
- [38] C. Milovic, B. Bilgic, B. Zhao, J. Acosta-Cabrero, C. Tejos, Fast nonlinear susceptibility inversion with variational regularization, *Magn. Reson. Med.* 80 (2) (2018) 814–821, <https://doi.org/10.1002/mrm.27073>.
- [39] C. Milovic, C. Prieto, B. Bilgic, S. Uribe, J. Acosta-Cabrero, P. Irrazaval, C. Tejos, Comparison of parameter optimization methods for quantitative susceptibility mapping, *Magn. Reson. Med.* 85 (1) (2021) 480–494, <https://doi.org/10.1002/mrm.28435>.
- [40] Y. Xiao, S. Bériault, G.B. Pike, D.L. Collins, Multicontrast multi-echo FLASH MRI for targeting the subthalamic nucleus, *Magn. Reson. Med.* 30 (5) (2012) 627–640, <https://doi.org/10.1016/j.mri.2012.02.006>.
- [41] Y. Xiao, V. Fonov, S. Bériault, F.A. Subaie, M.M. Chakravarty, A.F. Sadikot, G. B. Pike, D.L. Collins, Multi-contrast unbiased MRI atlas of a Parkinson's disease population, *Int. J. Comput. Assist. Radiol. Surg.* 10 (3) (2015) 329–341, <https://doi.org/10.1007/s11548-014-1068-y>.
- [42] Y. Xiao, V. Fonov, Berialt S Chakravarty, F.A. Subaie, A. Sadikot, G.B. Pike, G. Berstrand, D.L. Collins, A dataset of multi-contrast population-averaged brain MRI atlases of a Parkinson's disease cohort, *Data Brief* 12 (2017) 370–379, <https://doi.org/10.1016/j.dib.2017.04.013>.
- [43] Penny W., Friston K., Ashburner J., Kiebel S., Nichols T. (2006). *Statistical parametric mapping: the analysis of function brain images*, first edition.
- [44] R. Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, 2013. Available from, <https://www.R-project.org/>.
- [45] H. Long, W. Zhu, L. Wei, J. Zhao, Iron homeostasis imbalance and ferroptosis in brain diseases, *MedComm.* 4 (4) (2020) e298, <https://doi.org/10.1002/mco2.298>, 2023 Jun 26PMID: 37377861; PMCID: PMC10292684.
- [46] Q. Wu, Q. Ren, J. Meng, W.J. Gao, Y.Z. Chang, Brain iron homeostasis and mental disorders, *Antioxidants* 12 (11) (2023 Nov 13) 1997, <https://doi.org/10.3390/antiox12111997>. PMID: 38001850; PMCID: PMC10669508.
- [47] Y. Uchida, H. Kan, K. Sakurai, Y. Horimoto, E. Hayashi, A. Iida, N. Okamura, K. Oishi, N. Matsukawa, APOE ε4 dose associates with increased brain iron and β-amyloid via blood-brain barrier dysfunction, *J. Neurol. Neurosurg. Psychiatry* (2022 Apr 28) 328519, <https://doi.org/10.1136/jnnp-2021-328519> jnnp-2021Epub ahead of print. PMID: 35483916.
- [48] A. Toll, L. Blanco-Hinojo, D. Bergé, X. Duran, I. Canosa, T. Legido, F. Marmol, V. Pérez-Solà, E. Fernández-Egea, A. Mané, Multidimensional predictors of negative symptoms in antipsychotic-naïve first-episode psychosis, *J. Psychiatry Neurosci.* 47 (1) (2022 Jan 19) E21–E31, <https://doi.org/10.1503/jpn.210138>. PMID: 35046133; PMCID: PMC8789336.
- [49] M.J. Cuesta, P. Lecumberri, L. Moreno-Izco, J.M. López-Ilundain, M. Ribeiro, T. Cabada, R. Lorente-Omeñaca, G. de Erausquin, G. García-Martí, J. Sanjuan, A. M. Sánchez-Torres, M. Gómez, V. Peralta, Motor abnormalities and basal ganglia in

- first-episode psychosis (FEP), *Psychol. Med.* 51 (10) (2021) 1625–1636, <https://doi.org/10.1017/S0033291720000343>. JulEpub 2020 Mar 2. PMID: 32114994.
- [50] S. Heckers, Neural models of schizophrenia, *Dialogues Clin. Neurosci.* 2 (3) (2000 Sep) 267–279, <https://doi.org/10.31887/DCNS.2000.2.3/sheckers>. PMID: 22033839; PMCID: PMC3181608.
- [51] D. Mier, F. Schirmbeck, G. Stoessel, et al., Reduced activity and connectivity of left amygdala in patients with schizophrenia treated with clozapine or olanzapine, *Eur. Arch. Psychiatry Clin. Neurosci.* 269 (8) (2019 Dec) 931–940, <https://doi.org/10.1007/s00406-018-0965-4>. Epub 2018 Dec 11. PMID: 30539230; PMCID: PMC6841919.
- [52] S. Mueller, D. Wang, R. Pan, D.J. Holt, H. Liu, Abnormalities in hemispheric specialization of caudate nucleus connectivity in schizophrenia, *JAMA Psychiatry* 72 (6) (2015 Jun) 552–560, <https://doi.org/10.1001/jamapsychiatry.2014.3176>. PMID: 25830688; PMCID: PMC4630217.
- [53] X. Huang, W. Pu, X. Li, A.J. Greenshaw, S.M. Dursun, Z. Xue, H. Liu, Z. Liu, Decreased left putamen and thalamus volume correlates with delusions in first-episode schizophrenia patients, *Front. Psychiatry* 8 (2017 Nov 20) 245, <https://doi.org/10.3389/fpsy.2017.00245>. PMID: 29209237; PMCID: PMC5702009.
- [54] O'Brien, Dropping highly collinear variables from a model: why it typically is not good idea, *Soc. Sci. Q* 98 (2017) 360–375.
- [55] M. Gregorich, S. Strohmaier, D. Dunker, G. Heinze, Regression with highly correlated predictors: variable omission is not the solution, *Int. J. Environ. Res. Public Health* 18 (8) (2021) 4259, <https://doi.org/10.3390/ijerph18084259>.
- [56] R. Sanfelici, D.B. Dwyer, L.A. Antonucci, N. Koutsouleris, Individualized diagnostic and prognostic models for patients with psychosis risk syndromes: a meta-analytic view on the state of the art, *Biol. Psychiatry* 88 (4) (2020) 349–360, <https://doi.org/10.1016/j.biopsych.2020.02.009>.
- [57] J. Chen, N.J. Gong, K.T. Chaim, M.C.G. Otaduy, C. Liu, Decompose quantitative susceptibility mapping (QSM) to sub-voxel diamagnetic and paramagnetic components based on gradient-echo MRI data, *Neuroimage* 242 (2021 Nov 15) 118477, <https://doi.org/10.1016/j.neuroimage.2021.118477>. Epub 2021 Aug 14. PMID: 34403742; PMCID: PMC8720043.