Extracting interpretable signatures of whole-brain dynamics through systematic comparison

Annie G. Bryant [1,✉], **Kevin Aquino** [1,2], **Linden Parkes** [3,4], **Alex Fornito** [4], **and Ben D. Fulcher** [1,✉]

[1]School of Physics, The University of Sydney, Camperdown, NSW, Australia
[2]Brain Key Incorporated, San Francisco, CA, USA
[3]Department of Psychiatry, Brain Health Institute, Rutgers University, Piscataway, NJ, USA
[4]Turner Institute for Brain & Mental Health, Monash University, VIC, Australia

Abstract

Despite a rich and interdisciplinary literature on time-series analysis, the brain's complex distributed dynamics are typically quantified using only a limited set of manually selected representations and statistics. This leaves open the possibility that alternative dynamical properties may outperform those reported for a given application. Here we address this limitation by introducing a systematic procedure to compare diverse, interpretable features of functional magnetic resonance imaging (fMRI) data, drawn from comprehensive algorithmic libraries to capture both dynamical properties of individual brain regions (intra-regional activity) and patterns of statistical dependence between regions (inter-regional functional coupling). We apply our data-driven approach to investigate alterations to dynamical structures in resting-state fMRI in participants with schizophrenia (SCZ), bipolar 1 disorder (BP), attention-deficit hyperactivity disorder (ADHD), and autism spectrum disorder (ASD). Our findings broadly support the use of linear time-series analysis techniques for fMRI-based case–control analysis, while also identifying novel informative dynamical structures that have previously received less attention. While some simple statistical representations of fMRI dynamics perform surprisingly well (e.g., properties within a single brain region), we found performance improvements through combining intra-regional properties with inter-regional coupling—consistent with distributed, multifaceted changes to fMRI dynamics in neuropsychiatric disorders. The data-driven approach here enables the systematic identification and interpretation of disorder-specific dynamical signatures, with applicability beyond neuroimaging to diverse scientific problems involving complex time-varying systems.

complex systems | functional neuroimaging | brain dynamics | neuropsychiatric disorders | functional connectivity

Correspondence: *annie.bryant@sydney.edu.au, ben.fulcher@sydney.edu.au*

1 Introduction

The world is replete with complex systems, in which individual components exhibit localized activity that dynamically and flexibly gives rise to emergent macro-scale behaviors—from consumers participating in the global economy [1] to birds migrating as part of a flock [2]. One such system of great biological interest is the brain, wherein individual neuronal populations yield local activity dynamics that collectively drive complex behaviors like decision-making and emotion [3, 4]. In the brain—as with other types of complex systems—there are myriad ways to represent different types of dynamical patterns, many of which involve discretizing the brain into 'regions' [5, 6]. Neuroimaging modalities like functional magnetic resonance imaging (fMRI) measure brain activity within voxels that are typically aggregated into brain regions, yielding a region-by-time multivariate time series (MTS). While an MTS can be analyzed using statistics derived from a vast theoretical literature on time-series analysis [7], only a limited set of methods are

typically used to summarize fMRI properties. For example, most resting-state fMRI studies have examined functional connectivity (FC) between pairs of brain regions by computing the Pearson correlation coefficient [8, 9]—a zero-lag cross-correlation that assumes a joint bivariate Gaussian distribution, which may not always hold with resting-state fMRI data [9, 10]. However, MTS representations of fMRI data can be summarized at the level of individual regions, coupling between pairs of regions across distributed networks, or higher-order interactions amongst multiple regions [11].

The choice of how to represent the rich dynamical structure in a brain-imaging dataset—e.g., intra-regional activity (properties of the fMRI signal time series for a given region) or inter-regional coupling (statistical dependence of the fMRI signal time series for two regions)—and which statistical properties, or 'features', to measure is typically made manually. Of the few studies to compare alternative time-series features for functional connectivity or (less commonly) intra-regional coupling, the scope and number of the examined statistics is limited [12–14]. This problem of which dynamical properties to capture from a complex dynamical system, and which algorithms might best capture them, is common to many data-driven studies of complex systems—leaving open the possibility that alternative ways of quantifying dynamical structures may be more appropriate (e.g., yield better performance, clearer interpretation, or computational efficiency) for a given problem at hand.

Addressing the challenge of tailoring an appropriate methodology to a given data-driven problem, 'highly comparative' feature sets have recently been developed to unify a comprehensive range of algorithms from across the time-series literature, enabling broad and systematic comparison of time-series features (that would previously have been impractical). For example, the *hctsa* library includes implementations of thousands of interdisciplinary univariate time-series features [15, 16], and the *pyspi* library includes hundreds of statistics of pairwise interactions [17]. These algorithmic libraries have since been applied to systematically compare feature-based representations of diverse complex systems, from arm movement during various types of exercise [17] to light curves of different types of stars [18]. Emerging applications of this highly comparative time-series feature approach to neuroimaging data have highlighted novel biomarkers that are particularly informative for a given application, such as distinguishing fMRI scans obtained at rest vs while watching a film [17] or comparing neurophysiological dynamics with cortical microarchitecture [19] or structural connectivity [20]. Importantly, features defined outside of the neuroimaging space (such as dynamic time warping, originally developed for spoken word recognition [21]) have been shown to usefully quantify patterns in fMRI dynamics [22–24], underscoring the benefit of examining a comprehensive and interdisciplinary feature set.

However, prior highly comparative analyses have focused on either intra-regional activity or inter-regional coupling separately, which fails to tackle the issue of which representation is optimal for a given problem. To address this gap, here we introduce a systematic method to evaluate multiple data-driven univariate time-series features, pairwise interaction statistics, and their combination, that is generally applicable to data-driven problems involving distributed time-varying systems. In the context of fMRI MTS, combining properties of intra-regional activity and inter-regional coupling is motivated by findings that the two representations synergistically improve classification performance across a variety of clinical settings including schizophrenia [25], Alzheimer's disease [26], nicotine addiction [27], and mental stress [28]. Our method is applied to resting-state fMRI for case–control classification of four neuropsychiatric disorders as an exemplary case study: schizophrenia (SCZ), bipolar I disorder (BP), attention-deficit hyperactivity disorder (ADHD), and autism spectrum disorder (ASD). Most work in this space has focused on FC based on the hypothesis that neuropsychiatric disorders arise from dysfunction across distributed networks rather than individual regions [29–31], though this hypothesis has yet to be evaluated systematically. While intra-regional activity has been primarily overlooked, it can yield whole-brain maps of localized disruption, aiding interpretability and opening the door for new types of questions that are inaccessible using pairwise FC, such as how a given region responds to the targeted stimulation of another region [32].

Neuroimaging-based diagnosis of psychiatric disorders is an area of great interest, as diagnosis is currently based on behavioral criteria [33] and is hindered by patient heterogeneity [34, 35] as well as inter-rater reliability issues [36, 37]. Open-source resting-state fMRI datasets spanning neuropsychiatric disorders like schizophrenia [38] and autism spectrum disorder [39, 40] have inspired the publication of increasingly complex classifiers like deep neural networks,

which employ intricate algorithmic architectures that can be challenging to interpret [41–43]. This is at odds with the core goals of neuroimaging-based classification studies, which include: (1) unpacking the neuroanatomical and functional underpinnings of a given disorder; and (2) developing a classification framework that aids the diagnosis of future patients [33, 44–47]. Instead, we used linear support vector machine (SVM) classifiers to combine features based on clear algorithms derived from interdisciplinary time-series literature with simpler classical machine learning models, aiming to clarify the types of underlying dynamical processes that distinguish brain activity across disorders [32, 48]. Broadly, insights into the dynamical processes that are altered can deepen understanding of mechanisms underlying disease pathogenesis and progression, potentially yielding clinically translatable biomarkers [49, 50]. The present study is accompanied by a repository that includes all code needed to reproduce analyses and visuals presented here [51].

2 Results

Starting with a resting-state blood oxygen level-dependent (BOLD) signal MTS, in the form of brain regions sampled over time as shown in Fig. 1A(i-ii), we quantified both intra-regional dynamics (Fig. 1A(iii)) and pairwise coupling (Fig. 1A(iv)) using interpretable time-series features derived from a rich interdisciplinary scientific literature. We compared the ability of five different ways of statistically representing localized intra-regional dynamical properties and/or inter-regional FC to classify cases from controls in each of four neuropsychiatric disorders, as schematically depicted in Fig. 1B and summarized in Table 1. To summarize intra-regional BOLD activity fluctuations, we computed 25 univariate time-series features that included the *catch22* feature set [53], as schematically depicted in Fig. 1C. The *catch22* features were distilled from an initial library of over 7000 candidate features [16] to concisely capture diverse properties of local dynamics, such as distributional shape, linear and nonlinear autocorrelation, and fluctuation analysis ([53], further described in Sec. 4.2.1). We also included three additional features: the mean and standard deviation (SD)—based on our previous findings that these first two moments of the BOLD time series distributions are informative on their own [54]—and fractional amplitude of low-frequency fluctuations (fALFF) as a benchmark statistic for localized dynamics [55, 56]. Inter-regional FC was summarized using 14 statistics for pairwise interactions (SPIs) derived as a representative subset from an initial library of over 200 candidate SPIs in the *pyspi* library [17], including the Pearson correlation coefficient, as shown in Fig. 1C. This set of SPIs includes statistics derived from causal inference, information theory, and spectral methods, which collectively measure a variety of coupling patterns (directed vs undirected, linear vs nonlinear, synchronous vs lagged) that paint a fuller picture of inter-regional communication patterns [12, 17, 22].

Five distinct representations of BOLD activity were systematically compared, as illustrated in Figs. 1D(i)–(v):

1. A_{region} [Fig. 1D(i)] represents each fMRI time series using a set of features capturing 25 dynamical properties of an individual brain region. Strong classification performance in this representation suggests localized, region-specific changes in neural activity in the corresponding disorder.

2. A_{feature} [Fig. 1D(ii)] represents each fMRI time series using a set of features that captures a single dynamical property from all brain regions (82 for SCZ, BP, and ADHD, and 48 for ASD). Strong classification performance in this representation suggests that a specific property of intra-regional resting activity is altered in at least one part of the brain in the corresponding disorder.

3. $A_{\text{uni_combo}}$ [Fig. 1D(iii)] represents each fMRI time series using a set of features combining the 25 time-series properties of each individual brain region. If this representation out-performs A_{region} and/or A_{feature} individually, that would suggest changes to different properties across different regions that could not be fully captured by either reduced representation.

4. A_{FC} [Fig. 1D(iv)] represents each fMRI time series using a set of features that capture all pairs of inter-regional coupling strengths computed using a single SPI. Strong classification performance in this representation suggests that a specific type of inter-regional coupling is altered amongst one or more brain region pairs in the corresponding disorder.

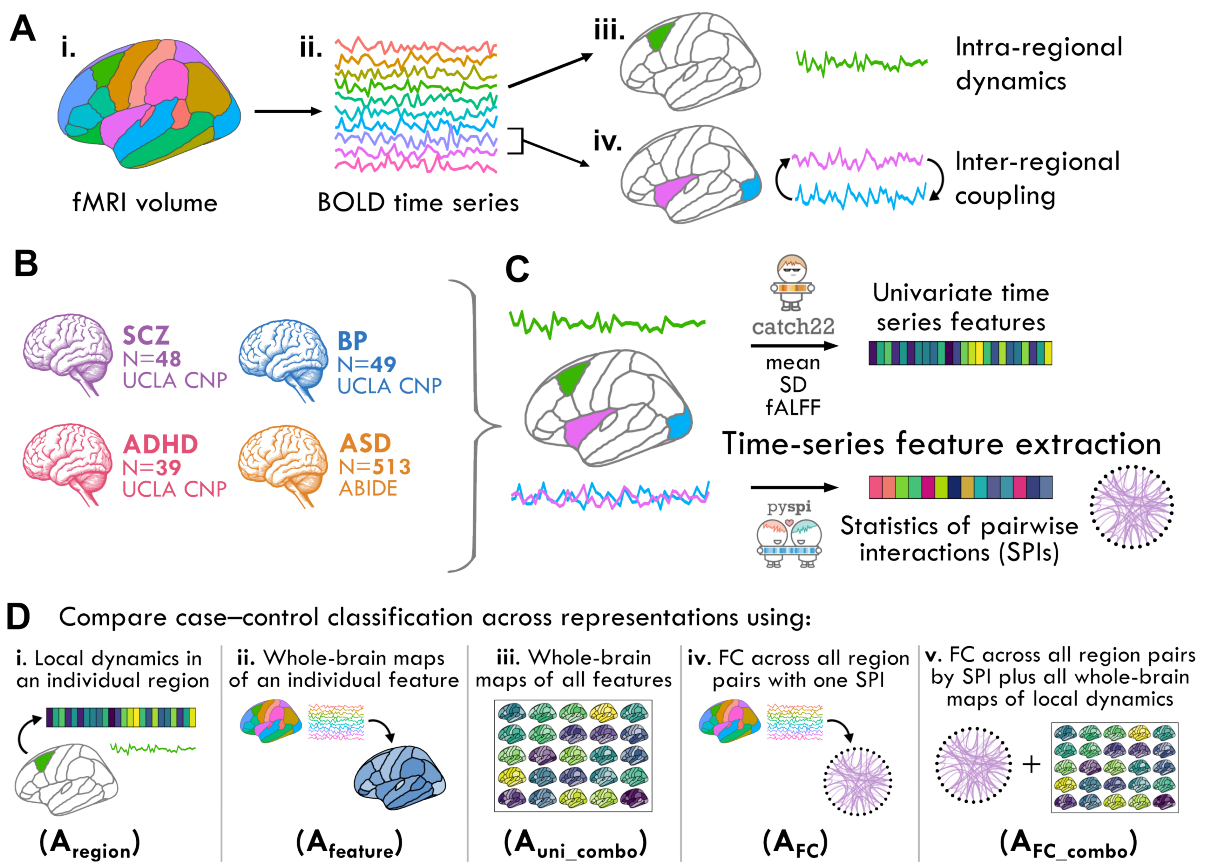

Figure 1. In this study, we systematically compare different ways of quantifying dynamical patterns from resting-state fMRI time series, focusing on statistics of local regional dynamics and pairwise coupling across four neuropsychiatric disorders. A For a given resting-state fMRI volume (**i**), the cortex and subcortex are parcellated into individual regions from which the voxel-averaged BOLD signal time series is extracted (**ii**). Two key ways of quantifying dynamical patterns from these data are to: (**iii**) measure properties of the dynamics of individual brain regions (shown green); or (**iv**) to compute statistical dependencies between pairs of regions (pink and blue). **B** To evaluate the performance of different types of dynamical representations of an fMRI time-series dataset (for identifying disease-relevant changes in neural activity), we included four neuropsychiatric exemplars originating from two open-access datasets: the UCLA CNP LA5c study [38] and the ABIDE I/II studies [39, 40, 52]. Each of the two cohorts also included cognitively healthy controls for comparison ($N = 116$ for UCLA CNP, $N = 578$ for ABIDE). **C** For each type of dynamical structure that we extracted from an fMRI dataset (i.e., for each feature-based 'representation' of the data), we computed interpretable time-series features that encapsulate a diverse range of activity properties. Local dynamical properties were quantified from each brain region using a set of 25 time-series features (the *catch22* feature set [53] along with the mean, SD, and fALFF). Interactions between all pairs of regions were quantified using a set of 14 statistics of pairwise interactions (SPIs) (a representative subset from the *pyspi* package [17]). Values for both univariate time-series features and sets of pairwise coupling strengths from SPIs were captured in one-dimensional feature vectors that encapsulate a given type of interpretable dynamical structure(s). **D** We evaluated case–control classification performance for each neuropsychiatric disorder using linear support vector machine classifiers fit to each of five distinct ways of representing resting-state fMRI properties: (i) all 25 univariate time-series features in an individual region, A_{region}; (ii) whole-brain maps of an individual time-series feature, $A_{feature}$; (iii) whole-brain maps of all 25 univariate time-series features, A_{uni_combo}; (iv) functional connectivity (FC) networks across all region pairs using one SPI, A_{FC}; and (v) A_{FC} along with all 25 univariate time-series features computed from all brain regions (A_{uni_combo}), termed A_{FC_combo}.

5. A_{FC_combo} [Fig. 1D(v)] represents each fMRI dataset as a set of features that capture all pairs of inter-regional coupling from a given SPI (as A_{FC}) as well as 25 time-series properties of all individual brain regions

(as A_{uni_combo}). If this representation out-performs A_{FC} on its own, that would suggest the presence of complementary and disease-relevant information in both localized intra-regional activity and pairwise coupling that is not captured with FC exclusively.

In all analyses, classification performance was evaluated using cross-validated balanced accuracy measured from a series of linear support vector machine (SVM) classifiers (see Methods Sec. 4.3.1 for details). Each of these analytical representations is discussed in its own subsection of the results, with the ultimate aim of comparing across these representations for each neuropsychiatric disorder.

2.1 Analysis A_{region}: Characterizing localized changes to intra-regional dynamics across disorders

Analysis A_{region} reduces the complexity of an individual's BOLD fMRI data down to a set of 25 time-series features that encapsulate the dynamical properties of an individual brain region, as schematically depicted in Fig. 2A. If neural activity in a single brain region is consistently disrupted across patients relative to controls, that will yield strong case–control classification performance; by applying this analysis to every single brain region individually, we can interpret the results through a spatial map of disrupted fMRI dynamics in each disorder. In this analysis, we test whether intra-regional dynamical properties of BOLD activity can meaningfully distinguish cases from controls in each of the four disorders—reasoning that disease-related changes to neural activity of a single brain region would yield strong case–control classification performance, which can be interpreted through a spatial map of disrupted fMRI dynamics across disorders. Such a pattern would contrast with previous reports that neuropsychiatric disorders like schizophrenia and ADHD are characterized by inter-regional dysfunction across distributed networks, rather than intra-regional disruptions [29, 55, 57, 58].

Multiple brain regions exhibited dynamics that significantly distinguished cases from controls in all four disorders (defined as Holm–Bonferroni corrected $p_{\mathrm{corr}} < 0.05$): $58/82$ individual regions in SCZ, $45/82$ regions in BP, $18/82$ regions in ADHD, and $27/48$ regions in ASD, with balanced accuracies for these significant regions listed in Supplementary Table S3. Across the four clinical groups, the brain region with the highest balanced accuracy varied markedly, from 54.9% in ASD up to 72.4% in SCZ as shown in Fig. 2B—noting that the threshold for significance varied across the four disorders due to differing sample sizes (see Methods Sec. 4.3.3 for details and Fig. S9 for null distributions). The spatial distribution of intra-regional differences in underlying dynamics (assessed via cross-validated balanced accuracy) is shown for each brain region in Figs 2C(i)–(iv) for SCZ, BP, ADHD, and ASD, respectively.

In SCZ, BOLD dynamics of medial occipital brain regions were the most informative of diagnosis, as shown in Fig. 2C(i). Specifically, against a background of nearly all brain regions yielding significant classification performance, dynamical signatures of activity in the left pericalcarine cortex best classified SCZ cases from controls, with $72.4 \pm 12.8\%$ balanced accuracy (permutation test $p_{\mathrm{corr}} = 9 \times 10^{-33}$, corrected across 82 brain regions), followed closely by the right cuneus ($70.0 \pm 11.8\%$, $p_{\mathrm{corr}} = 2 \times 10^{-27}$) and left cuneus ($69.2 \pm 11.5\%$, $p_{\mathrm{corr}} = 9 \times 10^{-26}$). Resting-state activity features in the pericalcarine cortex did not significantly distinguish BP or ADHD cases from controls, and the cuneal cortex exhibited significant but weaker performance in ASD [$52.0 \pm 4.2\%$, $p_{\mathrm{corr}} = 1 \times 10^{-2}$, corrected across 48 brain regions; Fig. 2C(iv)], which may indicate an SCZ-specific signature of disrupted intra-regional activity in these medial occipital regions. We confirmed that the classification performance across brain regions was not biased by gray-matter volume differences between cases versus controls across disorders, as had been previously posited [25] (cf. Sec. 4.3.4, Fig. S2A)—additionally, there was no bias in overall region volume on classification performance (Fig. S2B).

Subcortical structures, especially the thalamus, exhibited intra-regional dynamics that significantly distinguished cases from controls in each of the disorders (noting that subcortical data were not available for the ASD dataset). For BP, the right hippocampus achieved the highest balanced accuracy of all regions (permutation test $65.6 \pm 11.4\%$, $p_{\mathrm{corr}} = 9 \times 10^{-18}$, corrected across 82 brain regions), followed closely by the left caudate ($64.9 \pm 13.1\%$, $p_{\mathrm{corr}} = 2 \times 10^{-15}$) and left thalamus ($64.4 \pm 12.6\%$, $p_{\mathrm{corr}} = 7 \times 10^{-14}$), as indicated with darker

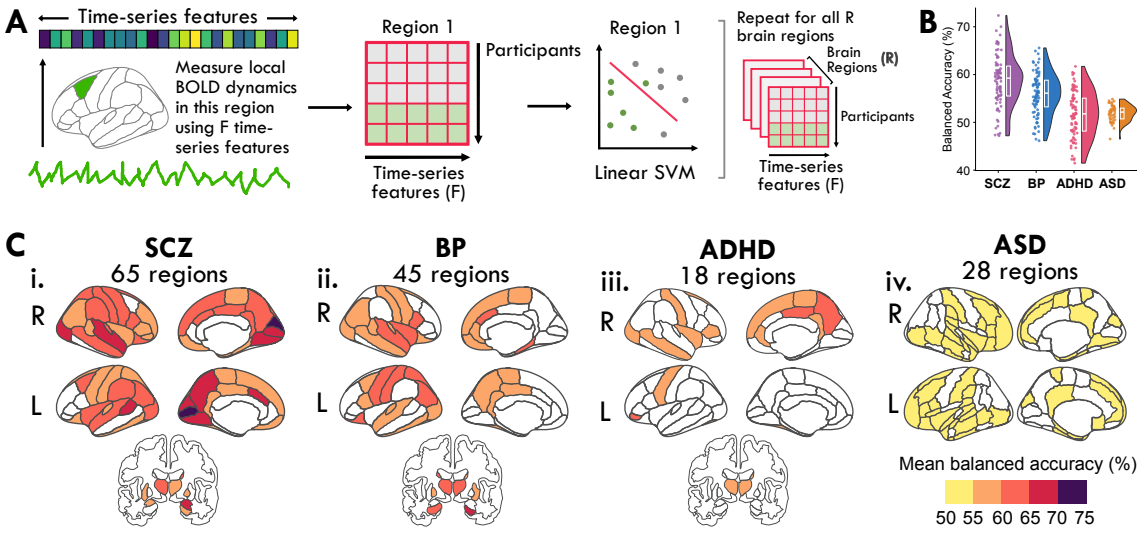

Figure 2. Local properties of BOLD dynamics in individual brain regions can significantly distinguish cases versus controls. **A** We consider the BOLD signal from an individual brain region (e.g., the caudal middle frontal gyrus shown in green) in which our set of 25 time-series features were computed. For each brain region, the values from all time-series features were organized into a feature $\times$ participant matrix, forming the basis for our case–control classification analysis at each individual region. **B** The balanced accuracy is shown for all brain regions as a raincloud distribution, where each point corresponds to the classification performance of a single region. The boxplot within each violin plot depicts the median and interquartile range for the region-wise classification performance within each neuropsychiatric disorder. **C** To visualize the spatial distribution of regional classification performance, mean case–control classification accuracies are depicted in brain maps for each disorder. Only regions that yielded a statistically significant balanced accuracy ($p_{\text{corr}} < 0.05$, corrected across 82 brain regions for SCZ, BP, and ADHD and 48 brain regions for ASD) are shaded, and the total number of significant brain regions per disorder is indicated above each brain map. Note that the Desikan–Killiany cortical and subcortical atlas is shown for SCZ (i), BPD (ii), and ADHD (iii) while the Harvard–Oxford cortical atlas is shown for ASD (iv).

shading in Fig. 2C(ii). Dynamical signatures of activity in the bilateral thalamus also significantly distinguished cases from controls in SCZ and ADHD, as shown in Figs 2C(i),(iii).

Fewer brain regions exhibited distinctive changes in resting-state dynamics in ADHD (18 out of 82 regions), with overall lower classification accuracies [all below 62%; cf. Fig. 2C(iii) and Table S3]. It is worth noting that the ADHD sample ($N = 39$) was smaller than that of SCZ ($N = 48$), BP ($N = 49$), or ASD ($N = 513$), though inverse probability weighting still boosted classifier performance beyond predicting the majority class for the top-performing regions in ADHD (cf. Fig. S3A). For the larger ABIDE-based ASD sample [39, 40], we found that more than half of the examined brain regions (27 out of 48) yielded significantly distinguishable dynamics relative to controls, but classification performance was low overall—with all balanced accuracies below 55% (cf. Fig. 2C(iv) and Table S3).

Here, we found that the dynamical signatures of multiple individual brain regions were significantly informative of diagnosis in each neuropsychiatric disorder—emphasizing some regions that are traditionally associated with psychopathology while highlighting relatively under-examined regions for future investigation.

2.2 Analysis $A_{feature}$: Assessing changes in brain-wide maps of individual dynamical properties across disorders

In this section, we asked whether an individual time-series feature of intra-regional dynamics measured across all brain regions could meaningfully distinguish cases from controls. As schematically depicted in Fig. 3A, we fit a separate model for each time-series feature individually, computed across all brain regions. Strong classification

performance in this representation would suggest that the given property is altered in one or more brain regions across cases of the corresponding disorder, although this representation cannot capture the case where multiple intra-regional dynamical properties are disrupted in a spatially heterogeneous manner.

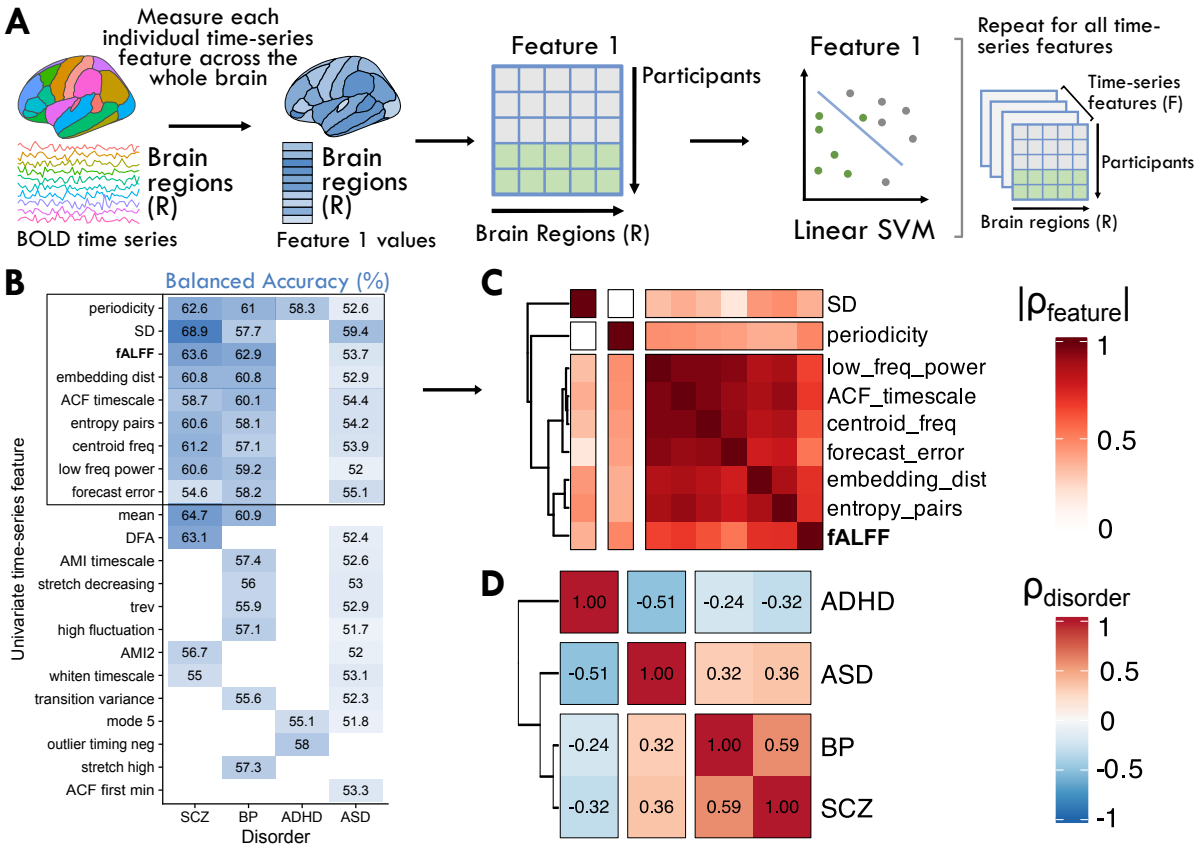

Figure 3. Analysis with individual time-series features measured across the brain highlights diverse differences in brain-wide dynamics across clinical groups. **A** For each of F (in this case, 25) time-series features, we examined its values computed from the BOLD signal across all brain regions in each disorder. For each univariate time-series feature, the values across all brain regions were organized into a region $\times$ participant matrix, which formed the basis for the case–control classification analysis. This process was repeated for each of the 25 features and for each disorder. **B** Classification results are shown as a heatmap, with rows representing univariate time-series features that yielded significant accuracy (permutation test $p_{\mathrm{corr}} < 0.05$, corrected across 25 features) in at least one disorder, and columns representing each of the four disorders. Of the 25 features evaluated, 22 significantly distinguished cases from controls in at least one disorder (these 22 features are shown here). A classical fMRI time-series metric, fALFF, is shown in boldface for reference. **C** The feature similarity score, $|\rho_{\mathrm{feature}}|$, between each pair of the 22 features from B is visualized as a heatmap, revealing three distinct clusters of features based on their values measured across all brain regions and all disorders. **D** The disorder similarity score, ρ_{disorder}, is depicted to compare the extent to which intra-regional features yielded similar classification performance between each pair of neuropsychiatric disorders. A large positive ρ_{disorder} value indicates a strong positive correlation in case–control classification performance across features in the given pair of disorders.

Brain-wide maps of 22 out of the 25 intra-regional time-series features yielded significant balanced accuracies in at least one of the four disorders (permutation test $p_{\mathrm{corr}} < 0.05$, corrected across 25 features), with balanced accuracies for these significant features listed in Supplementary Table S4. The mean classification performance of these time-series features is shown in Fig. 3B, with the maximum balanced accuracy ranging from $58.3 \pm 15.4\%$ in ADHD up to $68.9 \pm 13.2\%$ in SCZ—reflecting a lower range than that from looking at individual brain regions in A_{region} (Sec. 2.1). We found overall higher accuracies in SCZ and BP than in ADHD or ASD, reflecting a similar trend in the performance of models based on individual brain regions (A_{region}, see Fig. 2C). Only one time-series feature significantly distinguished cases from controls in all four disorders: a periodicity metric that measures

the first local maximum of the autocorrelation function (labeled as 'periodicity' here, as described in Wang et al. [59]). The fALFF (a commonly used fMRI time-series feature) also performed well, yielding significant classification performance for SCZ, BP, and ASD (though not for ADHD)—supporting its utility as a standard statistic for comparing disease-associated changes to local BOLD activity. None of the features achieved classification accuracies over 60% for ADHD, a threshold cleared by several individual brain regions in A_{region} (see Fig. 2B-C), suggesting that ADHD is not distinctively characterized by individual properties of intra-regional alterations to fMRI dynamics.

Across disorders, the other top-performing time-series features captured quantitatively similar properties to fALFF—namely, properties of linear autocorrelation and power spectrum structure. To confirm this, we focused on the nine time-series features that yielded significant accuracies in three or more disorders (outlined by a box in Fig. 3B) and computed a similarity index $|\rho_{\text{feature}}|$, defined as the absolute Spearman rank correlation between feature values measured from all brain regions and all participants. As shown in Fig. 3C, we found three distinctive clusters of features: one comprised of features capturing different aspects of linear autocorrelation structure, one comprised solely of 'periodicity', and one comprised solely of the standard deviation, SD. The high-performing linear autocorrelation-related properties included the fALFF and similarly related 'low_freq_power' (which measures the relative power in the lowest 20% of detected frequencies) as well as 'ACF_timescale', which suggests that properties related to the memory of the timescale (i.e., the extent to which past values influence current values of the BOLD signal) captured disease-relevant changes in fMRI dynamics. These linear features exhibited better classification performance than nonlinear properties like 'AMI_timescale'—the nonlinear analog to 'ACF_timescale', capturing the first minimum of the automutual information function—suggesting that linear properties were more sensitive to disease-related changes to fMRI dynamics in each disorder. Of note, while 'ACF_first_min' captured fairly similar properties to the fALFF ($|\rho_{\text{feature}}| = 0.57$) across all brain regions, it performed poorly overall, only significantly distinguishing ASD cases from controls, with low balanced accuracy ($53.3 \pm 4.3\%$, permutation test $p_{\text{corr}} = 2 \times 10^{-6}$, corrected across 25 features; see Fig. 3B). Altogether, we do not find compelling evidence that increasing the computational complexity beyond the fALFF is warranted for classification performance, suggesting that this linear and parsimonious feature is a suitable representative of this class of resting-state intra-regional dynamics over a comprehensive range of alternatives.

The BOLD SD and 'periodicity' exhibited more distinctive brain-wide values compared to the other cluster of linear properties, and also ranked among the highest-performing features across disorders, although they capture distinctive temporal properties to each other as evidenced by the $|\rho_{\text{feature}}|$ value of 0.09 (Fig. 3C). Notably, fALFF was less similar to 'periodicity' ($|\rho_{\text{feature}}| = 0.54$) than to the other top-performing linear autocorrelation features in the larger cluster in Fig. 3C (all $|\rho_{\text{feature}}| \geq 0.65$). This lower similarity suggests that periodicity captures a different aspect of intra-regional fMRI dynamics that is more distinctively altered in ADHD compared with fALFF—although with relatively low balanced accuracy ($58.3 \pm 15.4\%$, $p_{\text{corr}} = 2 \times 10^{-5}$). While BOLD SD has previously been reported as a powerful biomarker across case studies (including SCZ [60, 61] and ASD [62]; see *Discussion*), Power et al. [63] have also quantitatively linked brain-wide BOLD to head motion in the scanner, raising concern that head motion could be a confounding factor here. The two datasets (UCLA CNP and ABIDE) were preprocessed separately, which could yield differing residual motion artifacts [64], but we applied the same motion-based quality control filter to both datasets (see Methods, Sec. 4.1.3) such that we do not expect motion-related artefacts in the preprocessed BOLD time series to drive case–control differences. Still, we evaluated this potential confound by comparing the BOLD SD values averaged across all brain regions with the mean framewise displacement (FD) value per participant, which revealed significant whole-brain correlations for control participants in both studies (UCLA CNP and ABIDE) as well as ADHD and ASD participants (Fig. S4A). There was no association between BOLD SD and mean FD for SCZ (Spearman's $\rho = 0.19$, $p = 0.2$), which is notable as intra-regional BOLD SD was a top-performing feature for SCZ. Further examination within individual regions revealed significant associations between BOLD SD and mean FD for 11 brain regions in UCLA CNP and all 48 brain regions in ABIDE (cf. Fig. S4B). These associations indicate that BOLD SD is indeed related to head motion (as previously reported [65]), either reflecting incomplete nuisance signal regression and/or shared biological mechanisms driving neural signal variability and head movement.

In order to understand transdiagnostic similarities in the types of intra-regional properties that distinguish cases from controls, we computed a feature-based classification similarity score (ρ_{disorder}) for each pair of disorders

based on the Spearman rank correlation among classification accuracies for the 25 time-series features. A high ρ_{disorder} means that the set of 25 time-series features ranked similarly in classification performance between the two disorders. As shown in Fig. 3D, SCZ and BP exhibited the highest $\rho_{\text{disorder}} = 0.59$ ($p_{\text{corr}} = 0.01$, corrected across 6 pairwise comparisons), indicating that similar properties of intra-regional dynamics were informative of SCZ and BP diagnoses. Importantly, this does not imply that the features of localized activity were disrupted with a similar spatial topology between the two disorders, as SCZ and BP exhibited distinctive region-wise brain maps (in Fig. 2C). In ASD, performance across time-series features showed positive but weaker feature-based similarity scores with SCZ ($\rho_{\text{disorder}} = 0.36$, $p_{\text{corr}} = 0.3$) and BP ($\rho_{\text{disorder}} = 0.32$, $p_{\text{corr}} = 0.4$), suggesting modest similarity in the types of intra-regional features that were informative across these disorder comparisons. By contrast, ADHD exhibited the most distinctive time-series feature performance patterns to the other three disorders, with a negative ρ_{disorder} in each case, suggesting that the highest-performing features in the other disorders were the lowest performing features for ADHD, and vice versa. However, since only three of the 25 features yielded significant classification performance for ADHD, the negative correlation estimates are based on considerable noise—and should therefore be interpreted with caution.

In this section, $A_{feature}$ revealed that the fALFF and similar statistics related to linear autocorrelation performed well across disorders compared to other more complex and/or nonlinear statistics, and that similar properties of resting-state fMRI dynamics were distinctively altered in SCZ and BP.

2.3 Analysis A_{uni_combo}: Combining multiple brain regions and temporal properties to represent whole-brain fMRI dynamics across disorders

In contrast to the previous two reduced representations of intra-regional dynamics across the brain—in either a single region (A_{region}) or of a single time-series property ($A_{feature}$)—we next asked if a more comprehensive representation that integrates multiple time-series properties of resting-state activity measured from all brain regions could better distinguish cases from controls. An improvement in performance using this much higher-dimensional representation of the brain-wide dynamics (2050 dimensions for the UCLA CNP cohort: SCZ, BP, and ADHD, and 1200 dimensions for the ABIDE ASD cohort) would suggest that multiple properties of intra-regional activity measured across multiple spatial locations are important for characterizing a given disorder, in a way that cannot be simply reduced to a single region or single time-series feature. To investigate this question, we concatenated all intra-regional features computed from the whole brain into a single 'combination' classifier model, A_{uni_combo}, as depicted schematically in Fig. 4A.

Fig. 4B shows the mean balanced accuracy using this comprehensive representation of local dynamical properties across all brain regions as singular points in the middle row (A_{uni_combo}) for each disorder. We found that this representation distinguished cases from controls well for SCZ [$67.7 \pm 11.2\%$, permutation test $p = 4 \times 10^{-39}$, Fig. 4B(i)] and BP [$68.4 \pm 11.4\%$, $p = 2 \times 10^{-44}$, Fig. 4B(ii)], as indicated by the points located far from chance performance (shown as dashed lines at 50%). The strong performance in SCZ and BP was consistent with the large number of regions and features that yielded significant classification balanced accuracies on their own (cf. Figs 2B-C, 3B). The combination classifier also significantly distinguished ASD cases from controls (balanced accuracy $53.6 \pm 4.7\%$ [$p = 1 \times 10^{-10}$, Fig. 4B(iv)], although the low performance is likely attributable to the heterogeneity of the large multi-site sample [66–68]. By contrast, the combined region $\times$ feature classifier did not distinguish ADHD patients from controls better than the null model [cf. Sec. 4.3.3; $49.8 \pm 8.7\%$, $p = 5 \times 10^{-1}$, Fig. 4B(iii)]. Combination classifier performance is summarized in Supplementary Table S5 (for which the balanced accuracy was significant).

Since this combination classifier involves a considerably higher feature-space dimensionality, we asked whether the increased input feature space improved classification performance—or, alternatively, if a more parsimonious, reduced representation (A_{region} and/or $A_{feature}$) is more suitable for a given disorder. In Fig. 4B, to compare how well individual brain regions and time-series features compared with this combination classifier, we also depict the balanced accuracy for each individual brain region (top row, A_{region}) and each individual time-series feature (bottom row, $A_{feature}$) as a series of points. Figures 4B(i)-(ii) show that the combination of multiple intra-regional time-series

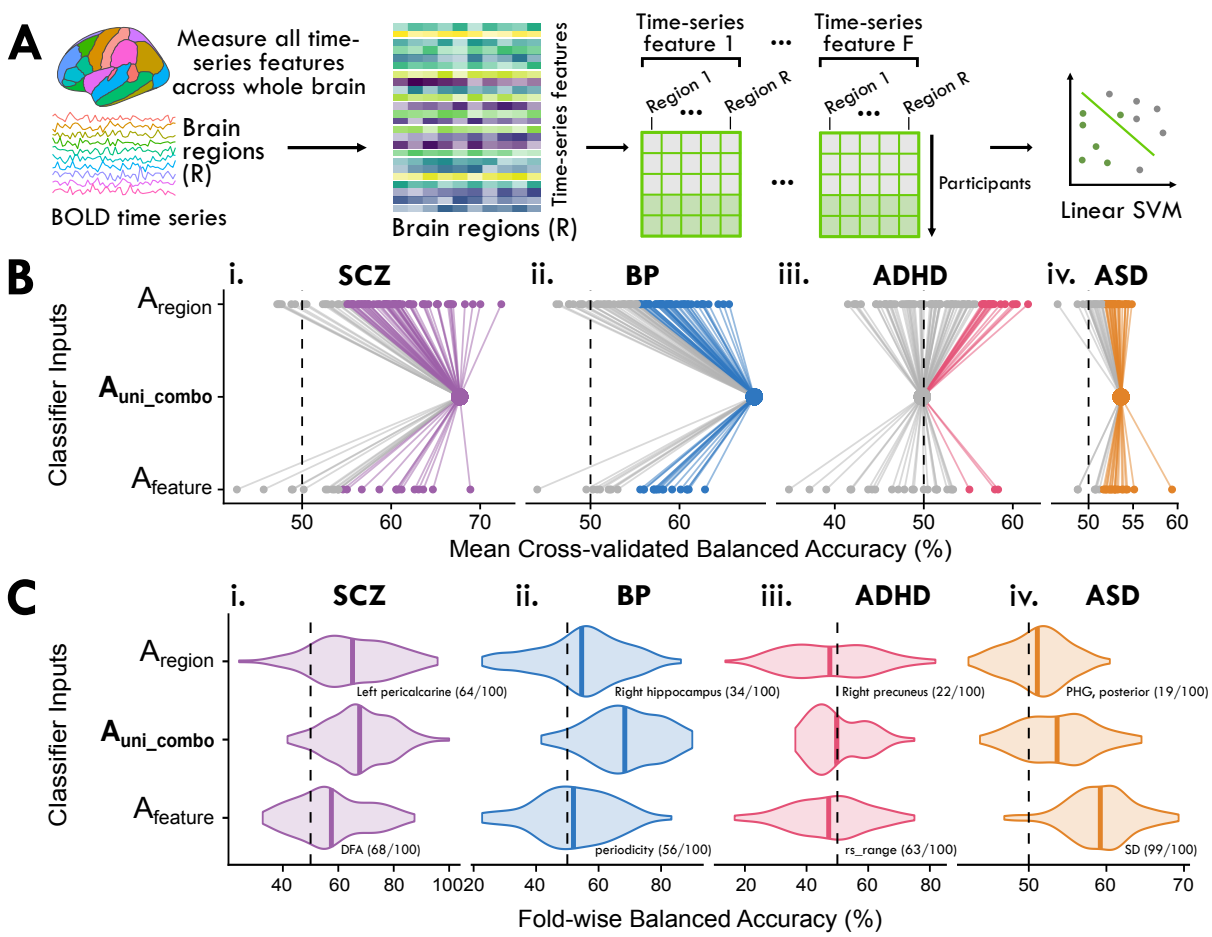

Figure 4. Representing a multivariate fMRI time series with whole-brain maps of univariate features improves classification performance for BP and SCZ, but not for ADHD or ASD—for which simpler representations perform better. A Starting with a multivariate resting-state BOLD time series, all F time-series features (in this case, 25 features) were computed across all R brain regions. All time-series feature values for all R brain regions were concatenated into one wide matrix, upon which the case–control classifier is evaluated for each disorder. **B** To compare the performance of this combination classifier with that of individual regions and time-series features, the mean balanced accuracy is shown for all brain regions (upper row, labeled A_{region}), the combination classifier (middle row, labeled A_{uni_combo}), and time-series features (bottom row, labeled $A_{feature}$). Lines are included as a visual guide to aid comparison. Colored dots indicate a classification model with significant balanced accuracy ($p_{\text{corr}} < 0.05$); gray dots indicate non-significant models. **C** To more directly compare performance across A_{region}, $A_{feature}$, and A_{uni_combo} representations of whole-brain univariate dynamical properties (which each involve fitting different numbers of models), we implemented a strategy that either selects the top-performing brain region (A_{region}) or the top-performing time-series feature ($A_{feature}$) in each training fold, and evaluated the distribution of balanced accuracy across test folds, shown as violin plots here. For each disorder, the brain region and time-series feature that were most frequently selected based on in-sample performance are annotated, along with the number of test folds out of 100 for which it was selected. Each vertical line indicates the mean out-of-sample balanced accuracy across the 100 train–test splits.

properties across multiple brain regions outperformed most individual-region (A_{region}) and individual-feature ($A_{feature}$) models in SCZ, and all such individual models in BP. By contrast, in ASD, multiple individual brain regions and time-series features surpassed the performance of the combination classifier [(Fig. 4B(iv)].

However, it is challenging to compare these different representation types fairly due to the different number of models evaluated across representations, i.e., one A_{uni_combo} model, 82 A_{region} models, and 25 $A_{feature}$ models for each disorder. To address this, for each of the 100 test folds (10 folds resampled 10 times), we identified the top-performing

brain region (for A_{region}) or time-series feature (for $A_{feature}$) based on each training fold (in-sample) balanced accuracy and evaluated its corresponding test (out-of-sample) balanced accuracy. This yielded a distribution of fold-wise balanced accuracy estimates based on a single test-fold prediction model for each representation—A_{region}, $A_{feature}$, and A_{uni_combo}—allowing us to directly compare the performance of each way of quantifying local dynamical structure across the brain. As shown in Figs 4C(i),(ii), on average, the combination classifier, A_{uni_combo}, better distinguished SCZ and BP cases from controls than using a lower-dimensional model based on a subset of data from the best individual brain region (A_{region}) or best individual time-series feature ($A_{feature}$). For ADHD, while there were many brain regions and time-series features that outperformed the combination classifier overall [Fig. 4B(iii)], selecting a single best feature or best region within each training portion did not yield high out-of-sample performance, as evidenced by the null out-of-sample balanced accuracy shown in Fig. 4C(iii). This strategy for selecting reduced feature subsets still yielded higher performance with the singular time-series feature representation ($A_{feature}$) for ASD [(Fig. 4C(iv)]—driven primarily by the BOLD SD, which was the feature selected in 99 out of 100 cross-validation folds.

For ADHD, there were fewer individual brain regions (18) and time-series features (3) that significantly distinguished cases from controls compared with the other disorders. We sought to test the possibility that the null performance of the A_{uni_combo} classifier may be at least partially attributable to noise accumulation [69]. As a simple test for overfitting via noise accumulation, we applied principal components analysis (PCA) to the full region $\times$ feature matrix for each disorder and the corresponding control cohort (see Fig. S5A for plots of the first two principal components). Since classification models based on individual brain regions yielded the lowest-dimensional classifiers (A_{region}: 25 time-series features per model), we selected the first 25 principal components (PCs) from the region $\times$ feature PCA for each case–control comparison group for consistency. Using the same classification procedure as implemented for the full region $\times$ feature matrix, we found that classifier performance with 25 PCs was comparable to that of the full feature space across all four disorders (cf. Fig. S5B). To complement this dimensionality reduction technique, we also evaluated the performance of L1-regularized SVM classifiers ('LASSO', see Methods Sec. 4.3.1), finding similar or poorer results with little evidence for classification improvement (cf. Fig. S5C). While a more exhaustive search for the optimal SVM hyperparameters and/or number of PCs might improve classifier performance, our preliminary analysis suggests that the poor performance of the combination classifier in ADHD is not due to overfitting alone, but rather to high heterogeneity in a small sample size.

2.4 Analysis A_{FC}: Comparing features of inter-regional functional connectivity across disorders

In this section, we examined the performance of classification models based on inter-regional coupling strengths between all pairs of regions, with the aim of understanding which statistical formulations of time-series dependencies are optimally suited to capture case–control differences in brain-wide FC in each of the four disorders. We compared 14 SPIs from the *pyspi* library [17] that capture different types of inter-regional coupling, including nonlinear, time-lagged, and frequency-based interactions, derived from literature including causal inference, information theory, and spectral methods (see Methods, Sec. 4.2.2 for more details).

As depicted schematically in Fig. 5A, each SPI was computed between the BOLD time series of all region–region pairs per participant and evaluated with the same cross-validated classification procedure (see Methods, Sec. 4.3.1). We found that 11 out of the 14 SPIs yielded a statistically significant balanced accuracy in at least one of the neuropsychiatric disorders, which are shown in a heatmap in Fig. 5B and summarized in Supplementary Table S6. The significant performance of multiple SPIs suggests that inter-regional communication patterns were disrupted across disorders in a way that pairwise FC metrics are generally well suited to capture, consistent with prior work [70, 71]. The maximum performance of any individual SPI ranged from $54.0 \pm 8.3\%$ in ADHD up to $69.1 \pm 12.0\%$ in SCZ, reflecting our findings in previous sections that SCZ was the most clearly distinguishable from controls while ADHD was the least distinguishable. Amongst the eleven significant SPIs, the Pearson correlation coefficient performed well, yielding some of the highest classification accuracies for SCZ ($69.1 \pm 12.0\%$, permutation test $p_{\text{corr}} = 2 \times 10^{-35}$, corrected across 14 SPIs), BP ($59.9 \pm 12.0\%$, $p_{\text{corr}} = 2 \times 10^{-11}$), and ASD ($59.1 \pm 3.9\%$, $p_{\text{corr}} = 3 \times 10^{-60}$). The strong performance of the Pearson correlation coefficient—which captures linear, synchronous

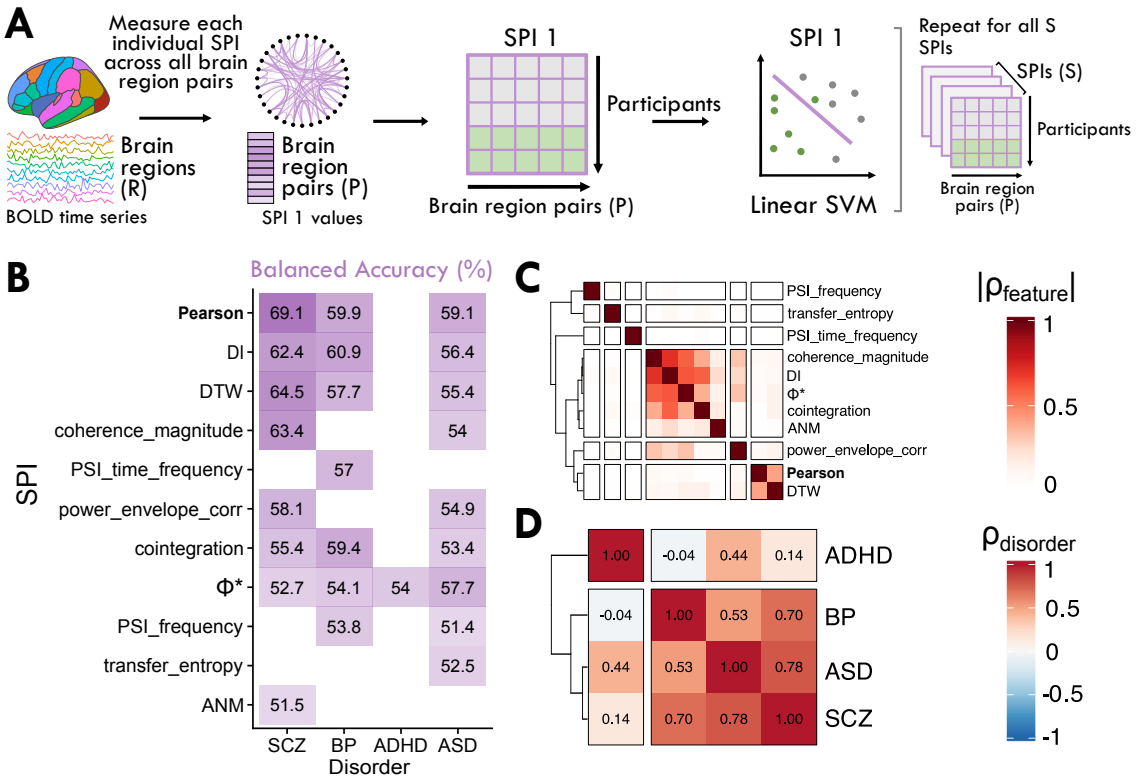

Figure 5. Representing brain activity as the set of all pairwise functional connectivity strengths, A_{FC}, can significantly distinguish cases from controls, with the classical Pearson correlation coefficient (capturing linear contemporaneous coupling) a top performing metric. A We compared 14 statistics of pairwise interactions (SPIs) (from *pyspi* [17], cf. Sec. 4.2.2), as different ways of quantifying functional connectivity (FC) between pairs of brain regions. For a given SPI, each participant was represented by the set of corresponding FC values (calculated for each pair of brain regions), yielding a set of region–region pair values that can be stored as a one-dimensional vector per participant. These vectors were concatenated to yield a participant $\times$ region–pair matrix that formed the basis for case–control classification using a linear SVM. **B** Classification results are shown as a heatmap, with rows representing SPIs that yielded significant balanced accuracy ($p_{\text{corr}} < 0.05$, corrected across 14 SPIs) in at least one disorder, and columns representing each of the four disorders. Of the 14 SPIs we evaluated, eleven significantly distinguished cases from controls in at least one disorder, and are plotted here. The Pearson correlation coefficient is annotated in boldface for easier reference. **C** The SPI similarity score, $|\rho_{\text{SPI}}|$, is visualized between each pair of the eleven SPIs from B as a heatmap, revealing six clusters of SPIs with similar behavior on the dataset (based on their outputs across all region–pairs and all disorders). As in B, the Pearson correlation coefficient annotation is shown boldface. **D** The disorder similarity score, ρ_{disorder}, is depicted to compare the balanced accuracy values among all 14 SPIs between each pair of neuropsychiatric disorders; a large positive ρ_{disorder} indicates a strong positive Spearman correlation in case–control classification performance across the 14 SPIs in the given pair of disorders.

inter-regional coupling under a Gaussian assumption—confirms the validity of these simplifying approximations in capturing disorder-relevant pairwise dependence structures from resting-state fMRI.

Next, we investigated the relationships between top-performing SPIs to clarify the types of conceptually distinct pairwise interaction statistics that differentiate cases from controls in each disorder. For each pair of SPIs, we computed an SPI similarity index, $|\rho_{\text{SPI}}|$, based on the absolute Spearman rank correlation between the two FC matrices pooled across all participants. We applied hierarchical clustering to the $|\rho_{\text{SPI}}|$ values from the 11 SPIs depicted in Fig. 5B. Fig. 5C shows the resulting six groups of top-performing SPIs, one of which included the Pearson correlation and dynamic time warping (DTW, $|\rho_{SPI}| = 0.40$)—with the latter also a top-performing SPI in SCZ, BP, and ASD (Fig. 5B). Interestingly, directed information (DI, computed with a Gaussian density estimator,

pyspi label 'di_gaussian') distinguished cases from controls with comparable balanced accuracy to that of Pearson correlation coefficient and DTW, despite exhibiting quite dissimilar behavior (as evidenced by the low SPI similarity index values, $|\rho_{\mathrm{SPI}}| = 0.05$, for Pearson correlation coefficient, and $|\rho_{\mathrm{SPI}}| = 0.08$ for DTW, cf. Fig. 5C). DI is a directed SPI that yields an asymmetric FC matrix, unlike the undirected Pearson correlation coefficient and DTW, for which only the non-redundant (supradiagonal) FC matrix values are included for classification. Its relatively strong performance implicates nonlinear and directed FC alterations between pairs of brain regions in SCZ, BP, and ASD, in a spatially distinctive manner to that of Pearson correlation- and DTW-based FC (given the low SPI similarity indices). DI was part of a larger cluster of SPIs that included coherence magnitude and cointegration (Fig. 5C), which generally exhibited poorer classification performance across disorders, as shown in Fig. 5B. This cluster also included Φ^*—an estimate for the extent of integrated information [72, 73] between the corresponding pair of regions—which was the only SPI to significantly distinguish cases from controls across all four disorders, although all balanced accuracy averages were below 60%. Notably, Φ^* was also the only SPI to significantly distinguish ADHD cases from controls, at 54.0 ± 8.3% (permutation test $p_{\mathrm{corr}} = 2 \times 10^{-6}$, corrected across 14 SPIs). However, this only slightly surpassed the Pearson correlation coefficient in ADHD (which yielded $53.2 \pm 10.7\%$ balanced accuracy, and was not significant, $p_{\mathrm{corr}} = 0.09$). The other two clusters in Fig. 5C contained the phase slope index (PSI) calculated in the time–frequency domain on its own, with transfer entropy (TE) and the PSI calculated in the frequency domain grouped together. These three SPIs yielded among the lowest balanced accuracies of the 11 SPIs depicted in Fig. 5B.

In order to understand whether similar FC metrics distinguished cases from controls well across disorders, we computed a disorder similarity index, ρ_{disorder}, that measured the Spearman rank correlation between all 14 SPIs for the given pair of disorders—noting that we included all SPIs in this analysis for completeness. As shown in the heatmap in Fig. 5D, the relative performance of SPIs was very similar between the three disorders for which the A_{FC} models showed significant performance: SCZ, ASD, and BP (all $\rho_{\mathrm{disorder}} > 0.5$). This suggests high similarity in the types of inter-regional coupling alterations that are relevant to these disorders. By contrast, there were minimal commonalities between ADHD and the other disorders, as expected given the null performance for all SPIs other than Φ^*.

Overall, in this section, we found that several FC statistics measured across all region–region pairs distinguished cases from controls with significant balanced accuracy for SCZ, BP, and ASD. For these three disorders (SCZ, BP, and ASD), FC statistics showed largely similar classification performance, with the Pearson correlation coefficient yielding among the highest balanced accuracy.

2.5 Analysis A_{FC_combo}: Integrating local regional dynamics with pairwise coupling into one classification model across disorders

In previous sections, we represented participant-level BOLD fMRI data using either properties of intra-regional dynamics or inter-regional FC separately. We finally asked whether a unified representation, that combines both local dynamics and pairwise coupling, would offer complementary information about resting-state fMRI activity that could better distinguish cases from controls. To test this, for each SPI, we concatenated the pairwise region–region matrix (A_{FC}) with the full region $\times$ univariate feature matrix (A_{uni_combo}) which served as classifier inputs for an A_{FC_combo} representation of an fMRI dataset, as depicted schematically in Fig. 6A.

All 14 SPIs significantly distinguished SCZ, BP, and ASD cases from controls with the inclusion of localized intra-regional dynamics, with balanced accuracies for each A_{FC_combo} model displayed in a heatmap in Fig. 6B and summarized in Supplementary Table S7. None of the A_{FC_combo} models were significant for ADHD. The peak classification performance across A_{FC_combo} models for a given disorder range from 59.4% in ASD to 71.6% in SCZ. As with our analyses based on pairwise FC alone (A_{FC}, cf. Fig. 5B), the Pearson correlation remained a top-performing SPI. DTW and DI were also among the top SPIs for SCZ, BP, and ASD, but with slightly lower performance than the Pearson correlation coefficient across the disorders. With the inclusion of intra-regional properties, Φ^* remained a top-performing SPI for ASD and rose to become a top-performing SPI for both SCZ and BP. While Φ^* was the only SPI to significantly distinguish ADHD cases from controls on its own (in A_{FC}), its

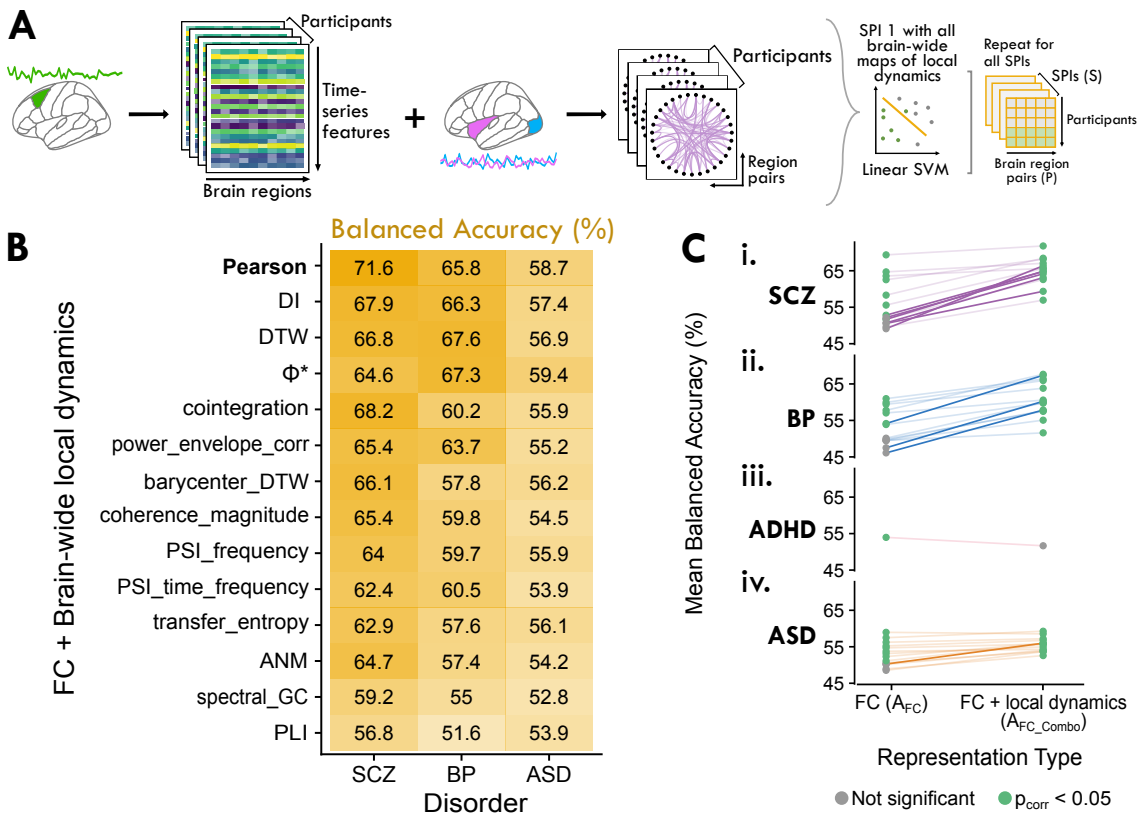

Figure 6. Adding statistics of local regional dynamics enhances classification performance for many pairwise FC metrics. A Here we evaluate the effect of including brain-wide maps of local regional dynamics with pairwise coupling data for each SPI on case–control classification performance. By concatenating the full univariate region $\times$ feature matrix from A_{uni_combo} with the full set of brain region pairs per SPI from A_{FC}, this yielded a set of 14 linear SVM classifiers. **B** The mean balanced accuracy is shown in heat map tiles for each SPI by row, organized into columns by neuropsychiatric disorder. ADHD is not included as none of the A_{uni_combo} models yielded a significant balanced accuracy. **C** The mean balanced accuracy is shown for each SPI on its own (left, A_{FC}) and with the inclusion of univariate region $\times$ feature data (right, A_{FC_combo}). Points are colored to indicate whether the corresponding balanced accuracy was significant (green, $p_{\mathrm{corr}} < 0.05$) or not significant (gray) relative to the corresponding null distribution (cf. Sec. 4.3.3). Each line maps to one SPI to visually guide comparison across representation types. Lines are shaded darker to indicate a significant difference in SPI performance with the inclusion of localized dynamics (corrected t-test $p_{\mathrm{corr}} < 0.05$, corrected across 14 SPIs; cf. Methods, Sec. 4.3.5).

performance dropped to null with the addition of brain-wide local dynamics (A_{FC_combo})—with the drop in performance in a higher-dimensional space likely arising from the small, heterogeneous sample.

We next aimed to more clearly understand whether the inclusion of univariate dynamics (A_{uni_combo}) into an SPI-based model (A_{FC}, based on pairwise coupling strengths) increased the model accuracy (A_{FC_combo}), despite involving a considerable increase in feature-space dimensionality. We compared the classification performance for each SPI on its own in the FC-based model, A_{FC}, to that of the combination model, A_{FC_combo}. The SPI-wise balanced accuracy for each type of model is depicted in Fig. 6C, revealing that most SPIs better distinguished cases from controls with the inclusion of brain-wide statistics of intra-regional dynamical properties. We quantified the difference in performance for each SPI using corrected t-tests ([74], see Methods Sec. 4.3.5), revealing six significant SPI improvements in SCZ ('barycenter_DTW', 'ANM', 'transfer_entropy', Φ^*, 'PSI_frequency', and 'spectral_GC'; Fig. 6(i)), three in BP ('power_envelope_corr', 'Φ^*', and 'barycenter_DTW'; Fig. 6(ii)), and one in ASD ('barycenter_DTW'; Fig. 6(iv); cf. Supplementary Table S8). Of particular interest is the significant improvement in Φ^* classification performance in BP, increasing from $54.1 \pm 8.0\%$ on its own, up to $67.3 \pm 11.7\%$ with the inclusion of

intra-regional dynamics ($t_{\mathrm{corr}} = 2.9, p_{\mathrm{corr}} = 0.03$, corrected across 14 SPIs). The addition of intra-regional time-series features did not just improve the performance of significant A_{FC} models, but also elevated the performance of non-significant SPIs to yield strong performance. This was the case with the barycenter (a measurement of the center of mass of the BOLD signal time series between the pair of brain regions, with the barycenter magnitude reflecting the extent of dynamic coupling [75, 76]), which showed significantly improved ability to distinguish cases from controls in all three disorders in A_{FC_combo} despite null performance on its own in A_{FC}.

3 Discussion

There are myriad ways in which a data analyst can extract interpretable feature-based representations of dynamical structures contained in a multivariate time series, like those measured with fMRI. From this statistical smorgasbord, a given researcher typically chooses a set of dynamical properties to study for a given problem subjectively, such as a set of pairwise Pearson correlation coupling strengths or the fALFF in a given set of brain regions. The lack of systematic methodological comparison thus makes it unclear from any given study whether these chosen dynamical properties are optimal, or whether alternative—and potentially simpler and more interpretable—statistics may outperform those reported. To address these concerns, here we introduced a systematic comparison of feature-based representations of fMRI time series, based on localized intra-regional dynamics, pairwise inter-regional coupling, and their combination, allowing us to systematically capture and compare different facets of fMRI dynamics. Our results demonstrate the benefits of such comparison by identifying the most parsimonious types of structures that are relevant for a given application, revealing disorder-specific signatures across four neuropsychiatric groups, particularly for SCZ and BP.

Our findings broadly support the use of linear time-series analysis techniques that are commonly used for fMRI data analysis, both at the level of individual brain regions (for which methods based on linear autocorrelation were top-performing metrics) and at the level of pairwise coupling (for which contemporaneous Pearson correlation was a top-performing metric), while simultaneously identifying novel high-performing metrics that warrant future investigation (including directed information to capture asymmetric information flows, and novel 'periodicity'-based univariate time-series features). While the intra-regional dynamical properties of some individual brain regions can yield good performance, we found improvement in combining dynamical properties across the whole brain (in SCZ and BP), consistent with the view that these neuropsychiatric disorders involve disruptions to brain dynamics that are spatially distributed and multifaceted [30, 50, 77]. Pairwise coupling strengths also served as an informative way to represent dynamical structures related to case–control differences, although our results demonstrated improvements when combining both intra-regional and inter-regional structures in a unified representation of an fMRI MTS dataset, again consistent with complex and spatially distributed disruptions. Our work provides a methodological foundation for systematically invoking representative features from a rich interdisciplinary literature on time-series analysis to determine the most appropriate way(s) of summarizing interpretable dynamical structures in MTS datasets. The approach presented here—which draws on recent algorithmic catalogs of the univariate and pairwise time-series analysis literatures [15, 17]—is generalizable to a wide range of problems across neuroimaging modalities and applications, as well as to a vast array of science and industry problems in which complex time-varying systems are measured and analyzed.

3.1 Multiple dynamical signatures of resting-state activity within individual brain regions are informative of diagnosis

Our results in A_{region} (in Results Sec. 2.1) demonstrate that statistical properties of univariate BOLD dynamics within many individual brain regions significantly distinguished neuropsychiatric patients from controls across the four disorders. The disorders differed substantially in the spatial distribution of the disruption—e.g., with widespread significant disruption in SCZ compared with relatively sparse disruption in ADHD—and in the distinguishability of these alterations—e.g., > 70% balanced accuracy in multiple regions in SCZ, but < 55% for all regions in ASD. In fMRI studies, intra-regional activity is often overlooked in favor of inter-regional functional connectivity, in part because of the hypothesis that neuropsychiatric disorders arise from disruptions to inter-regional coupling and

integration, with local dynamics alone insufficient to explain or predict diagnosis [25, 78, 79]. These conclusions about the lack of localized disruptions are largely based on previous work that has examined regional properties based on graph theory (e.g., degree centrality or node strength for a seed region in an FC matrix) [80, 81] or a handful of time-series features like the fALFF or regional homogeneity (ReHo) [82], which could miss more nuanced changes to the BOLD dynamics of a given brain region. A prevailing expectation is that fMRI is too noisy and poorly sampled in time for more sophisticated time-series analysis methods to provide additional useful information about intra-regional dynamics (e.g., that have been applied to EEG data) [83–85]. However, prior work comparing a large number of time-series features from individual brain regions using fMRI data has suggested that venturing beyond standard linear time-series analysis methods may be useful, such as the kurtosis-related distributional-shape features highlighted in Shafiei et al. [86]. Our results demonstrate that combining multiple properties of intra-regional fMRI activity from an individual brain region can meaningfully distinguish cases from controls in multiple neuropsychiatric disorders, suggesting that the breadth of features and/or the integration of multiple features into a unified statistical representation of fMRI time series can provide relevant information about disorder-specific alterations. Examining dynamics within individual brain regions enables spatial interpretability through visualizing brain-wide maps of classification performance, providing a clearly interpretable region-by-region picture of activity disruptions. This ability to characterize changes in BOLD dynamics at the level of individual regions has been key to addressing questions about regional differences in the response to spatially targeted brain stimulation [32, 87]. It also provides a clearer way to test molecular hypotheses about regional disruption in disorders which can be compared with rich multimodal region-level atlases spanning morphometry, cortical hierarchy, and multi-omics [86, 88, 89] to more deeply characterize the physiological underpinnings of disease-relevant changes in future work.

Across disorders, our findings from analysis A_{region} recapitulated prior results of individually informative brain regions, while also identifying novel region-specific changes in local dynamics that generate new hypotheses for future work. For example, resting-state properties of medial occipital brain regions like the pericalcarine and cuneus were particularly informative for SCZ classification, consistent with previous reports of altered fALFF in medial occipital regions [90–92] that may relate to disrupted perceptual organization [93] and/or visual hallucinations [94]. Additionally, regions in the subcortex like the thalamus and caudate nucleus exhibited resting-state dynamics that distinguished cases from controls across all three disorders (noting that subcortical data was not available for the ABIDE cohort). This is in line with prior studies highlighting subcortical disruption in BP [95, 96] and SCZ [97, 98], as well as compelling evidence that subcortical (specifically, striatal) dopamine dysfunction drives antipsychotic treatment efficacy [99, 100]. By contrast, subcortical alterations are seldom described in ADHD, in which pairwise FC receives the most attention [101].

Looking at individual properties of intra-regional dynamics measured across the whole brain in analysis A_{feature} (in Results Sec. 2.2), we found that properties related to linear autocorrelation (including Fourier power spectrum structure, such as fALFF) were top performers in distinguishing cases from controls across neuropsychiatric disorders (see Fig. 3C). The high classification metrics validate fALFF and related linear time-series analysis tools as useful techniques for capturing disorder-relevant dynamical properties of resting-state fMRI data. The conceptually related 'periodicity' feature (which captures the first peak in the autocorrelation function, satisfying conditions described in Wang et al. [59]) was the only statistic to significantly distinguish cases from controls in all four disorders. As this 'periodicity' has not been explored for fMRI previously (to our knowledge), future work with this statistic is warranted, with an opportunity to optimize some of its parameters for fMRI data specifically. The 'ACF_timescale' statistic was also a high-performing feature, which was highly correlated to the fALFF (as has been noted in prior work [20])—potentially detecting similar phenomena to the previously reported changes to intrinsic neural timescales in SCZ [102, 103] and ASD [103, 104]. Some of the time-series features we evaluated were conceptually similar to these top-performing linear autocorrelation metrics but yielded null classification accuracies, demonstrating that algorithms derived from theoretically similar foundations can display quite different performance in a given application. Interestingly, a nonlinear analog to the 'ACF_timescale'—'AMI_timescale', which captures the first minimum of the automutual information function (AMI)—performed worse than its linear counterpart across disorders. This is consistent with the view that BOLD dynamics (which are noisy and sparsely sampled in time) are well approximated by a linear stochastic process such that methods aiming to capture more complex (e.g., nonlinear)

dynamical structures may not be beneficial, as recently reported in Nozari et al. [105].

The BOLD SD stood out relative to the other top-performing univariate features, particularly as it is sensitive to the raw time-series values while all other features (besides the mean) were computed after z-score normalization (to focus on underlying dynamics of the time series in a way that does not depend on the measurement scale). We found that the brain-wide BOLD SD was particularly informative of SCZ and ASD diagnosis relative to controls (see Fig. 3B), recapitulating our previous work showing that the BOLD SD outperformed other univariate time-series features for SCZ [54]. Changes to BOLD signal variability have previously been described in both SCZ [60, 61] and ASD [62], with potential links to aberrant sensory integration. More broadly, regional BOLD SD alterations have been reported in settings ranging from healthy aging [106, 107] to Alzheimer's disease [108], and it is hypothesized that BOLD signal variance is linked to functional integration [109] as well as numerous molecular and cytoarchitectural properties [110]. Despite its widespread utility and reliability as a resting-state fMRI statistic [111], our observed association of SD with head movement (a non-neural confound) makes us cautious about interpreting BOLD SD-related changes as being driven by neural activity; this warrants further investigation in future work.

When we combined intra-regional dynamical properties measured across the entire brain in the combination model $A_{\text{uni_combo}}$ (in Results Sec. 2.3), we found that this comprehensive representation was particularly useful for classifying SCZ and BP cases relative to controls (see Figs 4B,C) The high classification performance of the combination classifier suggests that SCZ and BP are characterized by multiple types of changes to resting-state activity across multiple locations in the brain. Given the demonstrated biological heterogeneity in neuropsychiatric disorders (particularly SCZ) [112], effective exploration of larger time-series feature spaces could identify subtypes within a given disorder with distinctive dynamical profiles [113]—establishing the foundation for data-driven nosology [114]. By contrast to the expanded representations for SCZ and BP, reduced representations (i.e., individual brain regions or individual time-series features) better distinguished ADHD and ASD cases from controls than this combination classifier (see Fig. 3B), pointing to greater inter-individual heterogeneity relative to the (smaller) sample size for ADHD. In ASD, the BOLD SD was the dominant high-performing feature among the 25 we compared, emerging as the top-performing metric in 99 out of 100 training validations—which suggests that the other univariate time-series features contributed less disease-relevant information than the SD, demonstrating that systematic comparison can uncover potential model simplifications to select a parsimonious and interpretable feature-based representation.

3.2 Functional connectivity can better separate cases from controls with the inclusion of local dynamics

Through our comparison of SPIs in A_{FC} (see Results Sec. 2.4), we identified different ways of quantifying inter-regional coupling that were informative of case–control differences across the four psychiatric disorders. Out of the 14 evaluated SPIs, we found that the Pearson correlation coefficient was a top-performing statistic across disorders, particularly with the inclusion of whole-brain intra-regional dynamical properties in $A_{\text{FC_combo}}$ (see Results Sec. 2.5). Taken together with the strong case–control classification performance of fALFF measured across the whole brain, this suggests that linear time-series analysis methods are overall well-suited for capturing the salient dynamical properties of resting-state BOLD MTS [20, 86, 115]. Our comparisons also highlighted interesting alternative statistics with different behavior but similarly high performance to the Pearson correlation coefficient, such as DI [116, 117] and Φ^* [72, 73]—which involve more conceptually and computationally complex formulations of pairwise interactions that are seldom applied to fMRI datasets. While there was little evidence that increasing the complexity in an FC measurement beyond the Pearson correlation coefficient substantially improved neuropsychiatric disorder classification here, we previously found that alternative metrics like DI out-performed the Pearson correlation coefficient in an fMRI-based problem [17]. The alternative FC metrics examined here could be explored in future work using data with higher temporal resolution and signal-to-noise ratios (e.g., from MEG [19, 118])—a setting in which we might be able to detect more complex types of interactions in the brain (e.g., that are nonlinear and/or time-lagged). Of note, the Pearson correlation coefficient is undirected and therefore yielded half as many input attributes to the linear SVM compared with directed SPIs like DI (somewhat complicating a fair comparison between SPIs), in which case dimensionality reduction and/or feature selection techniques may be beneficial.

All 14 SPIs we evaluated in A_{FC_combo} (see Results Sec. 2.5) distinguished SCZ, BP, and ASD cases from controls with significant balanced accuracy when we also included brain-wide maps of all 25 intra-regional time-series features—with several significant improvements from the FC matrix on its own after correcting for repeated cross-validation and multiple comparisons. This finding supports previous work indicating that combining local and pairwise properties of resting-state fMRI data can synergistically improve classification performance [25–28].

3.3 SCZ and BP were more accurately distinguished from control participants than ADHD or ASD

It is difficult to directly compare balanced accuracy values across disorders given differences in sample sizes and acquisition sites (both of which are likely to attenuate out-of-sample classification performance [69]). However, we observed that, across all representations, SCZ and BP cases were distinguished from controls with higher balanced accuracy, while ASD and ADHD cases were more difficult to classify. By systematically comparing the same set of intra- and inter-regional time-series features across neuropsychiatric disorders, we quantified the degree of similarity in feature performance between pairs of disorders. SCZ and BP exhibited high similarity in the types of time-series features that were most informative of diagnosis, suggesting similar types of changes to resting-state fMRI dynamics—with previous work suggesting BP and SCZ share both symptom profiles and multi-omic alterations [119, 120].

While previous work suggests that ADHD also shares etiology with SCZ and BP based on self-reported personality traits and symptom domains [121], we found that ADHD was quite dissimilar to the other disorders in the types of features that distinguished cases from controls, particularly for localized intra-regional properties. We observed low classification performance in the ADHD group relative to SCZ and BP from the UCLA CNP cohort, which has previously been (partially) attributed to its smaller sample size [80]. However, the overall low classification performance contrasts with previous reports of higher classification performance in a smaller disorder class—which is partially due to various sources of bias [122, 123]. This inverse relationship between sample size and classification performance has also been reported in the ABIDE dataset previously [66–68, 124, 125]. There are a few potential explanations for why ADHD and ASD were less distinguishable from healthy controls than SCZ or BP in general. We included ASD and control participants from all ABIDE sites and did not account for imaging site in our classification analyses, following a recent meta-analysis reporting a lack of site-specific effects on classification performance [126]—although future work could evaluate the effect of multi-site integration techniques such as ComBat harmonization [127, 128] on intra- and inter-regional feature performance. Another difference with the ASD group relative to the other disorders is that homotopic region pairs were consolidated into one bilateral region, which could obscure hemisphere-specific changes to localized dynamics in ASD [129, 130]. It is possible that these populations exhibited changes to neural functional architecture that were more heterogeneous and individual-specific [123, 131, 132], making it difficult to identify an effective SVM decision boundary across representations. It is also possible that the ADHD and ASD samples generally included individuals with less severe symptoms than SCZ or BP cases [133], although we did not incorporate questionnaire responses or neuropsychological assessment data in the scope of this analysis. Alternatively, it could be that the features of intra- and inter-regional dynamics we examined were not sensitive to changes in fMRI dynamical structure that occur in ADHD and/or ASD.

3.4 Methodological limitations and considerations for future work

The findings presented here should be considered in the context of methodological limitations, described in following. While the goal of this study was to compare representations without optimizing classifier hyperparameters or performing feature selection, the balanced accuracy estimates reported here are lower than those reported in some other studies with the same datasets for SCZ [25, 134] and ASD [126]. As our main aim was to demonstrate how different types of interpretable dynamical structures could be compared systematically, we chose to use linear SVM classifiers to focus on simplicity. However, with such large feature spaces when multiple features are compared across thousands of brain-region pairs, alternative regularization may be beneficial—which could be accomplished through nested cross-validation for hyperparameter tuning and/or different regularization approaches,

such as graphical lasso [135]. The linear SVM effectively employed a Euclidean distance space, although Venkatesh et al. [136] note that (non-Euclidean) geodesic distance metrics are more suitable for comparing FC matrices across individuals. While we did not find performance improvements using SVMs with an RBF kernel or random forest classifiers (see Methods Sec. 4.3.1 for more details), it is possible that combining other nonlinear classifiers with hyperparameter optimization could more sensitively discriminate among features in a given disorder if nonlinear boundaries are present, warranting further study.

Several types of feature-based representations of fMRI time series—e.g., $A_{\text{uni_combo}}$, A_{FC}, and $A_{\text{FC_combo}}$—involved a high-dimensional feature space that exceeded the number of samples, a setting referred to as the '$p \gg n$' problem [137]. For example, while the combined region $\times$ feature matrix evaluated in the combination classifier $A_{\text{uni_combo}}$ encapsulated these reduced representations, increasing the number of input attributes to a classifier does not necessarily improve its performance—depending on the 'latent dimensionality' of the dataset, defined as the number of meaningful variables inferred from the data that capture underlying essential patterns [138, 139]. While the two dimensionality reduction techniques we evaluated (PCA and LASSO regularization) did not improve classification performance for ASD or ADHD, this was not an exhaustive investigation. Future work could systematically evaluate ways to reduce (spatial) dimensionality in a biologically informative manner (e.g., condensing region pairs into canonical functional networks [140] or applying similarity network fusion [141, 142]). It may also consider relaxing the picture of interacting spatially localized and contiguous brain regions, towards spatially distributed modes, extracted as components through dimensionality reduction [143, 144] or geometrically [145, 146].

Of note, previous work suggests that heterogeneous input data types (e.g., multiple brain regions, local dynamics versus pairwise coupling) require more nuanced data fusion techniques to be properly combined into one classifier [147]. Future applications of this systematic framework might consider such data fusion techniques, which are amenable to the inclusion of multimodal connectivity networks, as described in Hansen et al. [148], for example. While beyond the scope of this study, combining data fusion with other dimensionality reduction techniques can mitigate collinearity and enable accurate interpretation of linear SVM coefficients to clarify which brain regions/networks and time-series properties contribute more to a given decision boundary. More broadly, techniques to distill down the feature space will be of great utility to analyzing complex system MTS in general, both in terms of computational demand and mitigating noise accumulation.

Here we focused on five key ways of extracting intra-regional and inter-regional features from an fMRI dataset, but many other representations of spatiotemporal data could also be compared in future work. We supplied all edge weights between all regions directly as input attributes to the classifiers, but future work could reconfigure this representation to combine inter-regional graph network measures for a given seed region with its intra-regional dynamics. This includes quantifying properties of the networks defined by pairwise FC matrices (cf. highly comparative graph analysis [149]), statistics of spatiotemporal patterns (like spirals and traveling waves [150, 151]), and higher-order (beyond pairwise) dependence structures [152]. We also focused here on representative subsets of the univariate time-series analysis literature (the *catch22* subset of over 7000 features in the *hctsa* feature library [16]) and pairwise dependence literature (14 representative SPIs from the full *pyspi* library of over 200 SPIs [17]). Note that neither the univariate features nor the SPIs measured here were specifically tailored for neuroimaging applications; for example, the *catch22* set was derived based on classification performance across 128 diverse univariate time series datasets ranging from beef spectrograms to yoga poses [153]. Future work could consider ways to incorporate the full sets of local and pairwise time-series properties or to derive reduced subsets with algorithmic approaches tailored to a particular dataset. Broadening the scope of comparison in these ways—of both types of dynamical structures and algorithms for quantifying them—comes with associated issues of statistical power required to reliably pin down specific effects that future work will need to carefully consider.

4 Methods

An overview of our methodology is illustrated in Fig. 1. We extracted the BOLD time series from each resting-state fMRI volume, which were analyzed at the level of individual brain regions and pairs of brain regions (Fig. 1A). We

included participants from four neuropsychiatric disorder groups derived from two main cohort studies, as depicted in Fig. 1B and summarized in Table 1. For each participant, we computed a set of time-series features reflecting the local dynamics of a single region's BOLD time series and coupling between a pair of brain regions (Fig. 1C). We compared the performance of these univariate and pairwise features separately, and in combination, using linear classifiers for each disorder—specifically testing five different ways of capturing dynamical structure, as shown in Fig. 1D and outlined in the following.

4.1 Neuroimaging datasets and quality control

4.1.1 UCLA CNP

Raw BOLD resting-state fMRI volumes consisting of 152 frames acquired over 304 seconds were obtained from the University of California at Los Angeles (UCLA) Consortium for Neuropsychiatric Phenomics (CNP) LA5c Study [38]. This included cognitively healthy control participants as well as participants diagnosed with schizophrenia (SCZ), attention-deficit hyperactivity disorder (ADHD), and bipolar I disorder (BP). Details of diagnostic criteria and behavioral symptoms have been described previously, along with imaging acquisition details [38]. Imaging data were preprocessed using the *fmriprep* v1.1.1 software [154] and the independent component analysis-based automatic removal of motion artifacts (ICA–AROMA pipeline [155], as described previously [156]). Additional noise correction was performed to regress out white matter, cerebrospinal fluid, and global gray-matter signals using the ICA–AROMA + 2P + GMR method [156–158]. Time series of 152 samples in length were extracted from the noise-corrected BOLD resting-state fMRI volumes for 68 cortical regions [159] and 14 subcortical regions [160] spanning the left and right hemispheres.

4.1.2 ABIDE

Preprocessed BOLD resting-state fMRI time-series data consisting of 180 time points were obtained from the Autism Brain Imaging Data Exchange (ABIDE) I & II consortium [39, 40], working with a dataset specifically aggregated for an international classification challenge by Traut et al. [52] that included healthy controls and participants with autism spectrum disorder (ASD). Imaging data was preprocessed by the classification challenge authors [52, 126] using the standard pipeline from the '1000 Functional Connectomes Project', which includes motion correction, skull stripping, spatial smoothing, segmentation, and noise signal regression [126]. We opted to use the preprocessed time series corresponding to the Harvard–Oxford cortical atlas, which is published with the FSL software library [161] and includes 48 regions, noting that homotopic region pairs were consolidated into singular bilateral regions.

4.1.3 Quality control

After preprocessing, we identified six participants from the UCLA CNP dataset (IDs: sub-10227, sub-10235, sub-10460, sub-50004, sub-50077, and sub-70020) for whom the ICA–AROMA + 2P + GMR pipeline yielded constant zero values across all timepoints and brain regions. These six participants ($N = 3$ Control, $N = 2$ SCZ, $N = 1$ ADHD) were excluded from all further analyses. Head movement in the scanner was evaluated using framewise displacement (FD) calculated from the six framewise head-motion parameters (corresponding to rotation and translation in the x, y, and z planes) using the method of Power et al. [162]. We found that SCZ, BP, and ASD—but not ADHD—cases exhibited significantly higher mean FD relative to control participants, consistent with prior studies [163–165] (all $p < 0.05$, Wilcoxon rank-sum test). Head movement (mean FD) alone significantly distinguished cases from controls for SCZ ($74.0 \pm 12.5\%$), BP ($60.7 \pm 11.4\%$), and ASD ($55.5 \pm 4.3\%$), though not for ADHD ($44.0 \pm 11.5\%$; Fig. S6B). In order to mitigate the potential confounding effect of head motion on the dynamical BOLD properties with putative neural causes, we applied the movement-based exclusion criteria described in Parkes et al. [157] (labeled as 'lenient' criterion in that study), such that any participant with mean FD > 0.55 mm was excluded from further analysis. This step excluded two participants from the UCLA CNP study ($N = 1$ SCZ, $N = 1$ Control) and 59 participants from the ABIDE study ($N = 36$ ASD, $N = 23$ Control).

4.1.4 Participant summary statistics

After performing quality control, we retained between $N = 39$ and $N = 578$ participants per clinical group across the two studies, with summary statistics provided in Table 1. There were significant differences in age between controls and SCZ (Wilcoxon rank sum test, $p = 7 \times 10^{-4}$) and BP (Wilcoxon rank sum test, $p = 0.01$), but not between controls and ADHD or ASD. There were also significant differences in the distribution of males versus females between UCLA CNP controls and SCZ ($X^2(1, N = 164) = 5.7$, $p = 0.01$) and between ABIDE controls and ASD ($X^2(1, N = 1091) = 20.6$, $p = 5 \times 10^{-6}$), though not for BP or ADHD. Participants in the UCLA CNP dataset were scanned at two different sites [38] and participants in the ABIDE dataset were scanned across seventeen sites [126]. The ABIDE dataset is approximately ten-fold larger than any of the comparisons from the UCLA CNP dataset, and some studies have shown that classification performance with this dataset is inversely proportional to the number of participants included in the model [66–68]. As we did not include site information as a confounding variable in our classifiers, nor did we attempt to perform cross-site alignment, we expect that our classification results for the ABIDE dataset are lower than could be obtained using a singular site or by applying cross-site harmonization.

Table 1. Key demographics for participants included in this study after quality control exclusion processes.

Study	Diagnosis	*N*	% Female (*N*)	Age; Mean (SD)
ABIDE	ASD	513	14.6 (75)	17.5 (10.5)
	Control	578	26.0 (152)	17.1 (9.5)
UCLA CNP	ADHD	39	48.7 (19)	31.6 (10.1)
	BP	49	42.9 (21)	35.3 (9.0)
	SCZ	48	25.0 (12)	36.6 (9.0)
	Control	116	46.6 (54)	31.2 (8.7)

4.2 Time-series feature extraction

4.2.1 Univariate features

For a given brain region, the BOLD time series can be summarized using a set of univariate time-series features. We opted to use the *catch22* library of univariate properties [53], which are listed in Table 2 and described in more detail in Supplementary Table S1. The *catch22* features were drawn from the broader *hctsa* [15, 16] library, designed to be a highly explanatory subset of time-series features for generic time-series data mining problems [153] (though not defined specifically for neuroimaging datasets). To compare the performance of these features with a standard biomarker for quantifying localized resting-state BOLD activity, we also computed the fractional amplitude of low-frequency fluctuations (fALFF), which measures the ratio of spectral power in the 0.01–0.08 Hz range to that of the full frequency range [56]. fALFF is considered to be an index of spontaneous activity in individual brain regions and is fairly robust to the noise, scanner differences, and low sampling rates typical of resting-state fMRI data [166–168]. The *catch22* features were computed in R using the `theft` package (version 0.4.2; [169]) using the `calculate_features` function with the arguments `feature_set = "catch22"`, which z-scores each time series prior to feature calculation. We also calculated the mean and standard deviation (SD) from the raw BOLD time series by including the `catch24 = TRUE` argument, which yielded a total of 24 univariate time-series features. We computed fALFF in Matlab as per the code implementation included in Fallon et al. [20]. Collectively—the *catch22* features, mean, SD, and fALFF—comprise 25 univariate features that encapsulate a diverse set of statistical time-series properties with which to compare across different clinical exemplars. These 25 time-series features were computed across all brain regions and participants, yielding a three-dimensional array in the form of $N \times R \times F$—comprised of N participants, R brain regions, and F time-series features.

4.2.2 Pairwise features

To summarize different types of pairwise coupling, or functional connectivity (FC), across pairs of brain regions, we used a subset of 14 SPIs from the *pyspi* library of pairwise interaction statistics [17]. Each of these SPIs is

Table 2. The 25 univariate time-series features computed for each brain region to capture intra-regional dynamics. For each univariate time-series feature, its shorthand name (referenced throughout the manuscript and in figures) as well as its name in 'catch22' [53] code implementation is listed. More information about each feature beyond its code implementation name is provided in Table S1.

Shorthand name	***catch22/theft* code implementation**
embedding_dist	CO_Embed2_Dist_tau_d_expfit_meandiff
ACF_timescale	CO_f1ecac
ACF_first_min	CO_FirstMin_ac
AMI2	CO_HistogramAMI_even_2_5
trev	CO_trev_1_num
mode_10	DN_HistogramMode_10
mode_5	DN_HistogramMode_5
mean	DN_Mean
outlier_timing_neg	DN_OutlierInclude_n_001_mdrmd
outlier_timing_pos	DN_OutlierInclude_p_001_mdrmd
SD	DN_Spread_Std
whiten_timescale	FC_LocalSimple_mean1_tauresrat
forecast_error	FC_LocalSimple_mean3_stderr
AMI_timescale	IN_AutoMutualInfoStats_40_gaussian_fmmi
high_fluctuation	MD_hrv_classic_pnn40
periodicity	PD_PeriodicityWang_th0_01
stretch_decreasing	SB_BinaryStats_diff_longstretch0
stretch_high	SB_BinaryStats_mean_longstretch1
entropy_pairs	SB_MotifThree_quantile_hh
transition_variance	SB_TransitionMatrix_3ac_sumdiagcov
DFA	SC_FluctAnal_2_dfa_50_1_2_logi_prop_r1
rs_range	SC_FluctAnal_2_rsrangefit_50_1_logi_prop_r1
low_freq_power	SP_Summaries_welch_rect_area_5_1
centroid_freq	SP_Summaries_welch_rect_centroid
fALFF	fALFF

Table 3. The 14 statistics for pairwise interactions (SPIs) computed across brain region pairs. For each SPI, its shorthand name (referenced throughout the manuscript and in figures) as well as its name in 'pyspi' code implementation is listed. More information about each SPI beyond its code implementation name is provided in Table S2.

Shorthand name	***pyspi* code implementation**
PSI_frequency	psi_multitaper_mean_fs-1_fmin-0_fmax-0-5
ANM	anm
DI	di_gaussian
transfer_entropy	te_kraskov_NN-4_DCE_k-max-10_tau-max-4
phi_star	phi_star_t-1_norm-0
spectral_GC	sgc_nonparametric_mean_fs-1_fmin-0_fmax-0-5
PLI	pli_multitaper_max_fs-1_fmin-0_fmax-0-5
PSI_time_frequency	psi_wavelet_mean_fs-1_fmin-0_fmax-0-5_mean
barycenter_DTW	bary_dtw_mean
DTW	dtw_constraint-itakura
power_envelope_corr	pec
coherence_magnitude	cohmag_multitaper_mean_fs-1_fmin-0_fmax-0-5
cointegration	coint_aeg_tstat_trend-ct_autolag-aic_maxlag-10
Pearson	cov_EmpiricalCovariance

listed along with its code name in 'pyspi' in Table 3; more detailed descriptions are provided in Supplementary Table S2. These 14 SPIs contain a single representative from each of the 14 data-driven modules of similarly performing SPIs described in Cliff et al. [17]. It includes the most commonly used FC metric for fMRI time series, Pearson correlation coefficient (denoted as 'cov_EmpiricalCovariance' in *pyspi*), enabling direct comparison of its performance to the other thirteen evaluated SPIs, which span nonlinear coupling (e.g., additive noise modeling), frequency-based coupling (e.g., power envelope correlation), asynchronous coupling (e.g., coherence magnitude), time-lagged coupling (e.g., dynamic time warping), and directed coupling (e.g., directed information). The 14 SPIs were computed from the BOLD time series for each pair of brain regions using the *pyspi* package (version 0.4.0) in Python [170]. Given the large number of MTS matrices to process (across all participants), we distributed *pyspi* computations on a high-performance computing cluster. Each SPI yielded a matrix of pairwise interactions between all brain regions, with a total of 6642 possible pairs for UCLA CNP (from $R = 82$ brain regions) and 2256 pairs for ABIDE (from $R = 48$ brain regions), after omitting self-connections. As some SPIs (like the Pearson correlation) are undirected, the resulting FC matrix is symmetric, and only the set of coupling values for all unique brain-region pairs (corresponding to the upper half of the FC matrix) is retained for these SPIs. Across all SPIs, *pyspi* computations yielded a three-dimensional array in the form of $N \times P \times S$, comprised of N participants, P region pairs (noting that this value will differ for directed versus undirected SPIs), and S SPIs.

4.3 Case–control classification

We sought to comprehensively compare the ability of diverse intra-regional time-series features and inter-regional coupling strengths to separate healthy controls from individuals with a given clinical diagnosis. Univariate, intra-regional time-series features are sensitive to local, region-specific disruptions in neural activity, whereas inter-regional SPIs are sensitive to disrupted coupling between pairs of brain regions. We first asked whether case versus control separation would be better with: (i) multiple dynamical properties measured from a single brain region; (ii) a single time-series property measured across brain regions; or (iii) the combination (the full set of time-series properties across all brain regions). We then asked how pairwise FC measurements could capture case–control differences, and whether including our set of diverse univariate features would enhance SPI-wise classification performance. To compare the ability of univariate and/or pairwise features to separate case versus control individuals, we evaluated five different classification pipelines as enumerated in the following subsections, labeling each analysis as 'A_X' where X captures a short description of the analysis. These are listed as follows:

1. A_{region}: The performance of individual brain regions given all F univariate time-series features, assessed via classifiers fit to $N \times F$ (participant $\times$ feature) matrices as depicted in Fig. 1D(i). $R = 82$ models are fit with this representation for SCZ, BP, and ADHD and 48 models for ASD.

2. A_{feature}: The performance of individual univariate time-series features measured from all brain regions, assessed via classifiers fit to $N \times R$ (participant $\times$ region) matrices as depicted in Fig. 1D(ii). $F = 25$ such models are fit for all four disorders.

3. $A_{\text{uni_combo}}$: The performance of all F univariate time-series features computed across all R brain regions, assessed via classifiers fit to $N \times RF$ (participant $\times$ brain-region—feature) matrices as depicted in Fig. 1D(iii). Only one such model is fit per disorder.

4. A_{FC}: The performance of the set of coupling strengths measured from all brain region pairs using an individual SPI, assessed via classifiers fit to $N \times P$ (participant $\times$ region–pair) matrices, as depicted in Fig. 1D(iv). $S = 14$ such models are fit for each disorder.

5. $A_{\text{FC_combo}}$: The performance of an individual SPI measured from all brain region pairs with the addition of all F univariate time-series features, assessed via classifiers fit to $N \times PRF$ (participant $\times$ region-pair–region–univariate-feature) matrices as depicted in Fig. 1D(v). $S = 14$ such models are fit for each disorder.

4.3.1 Classifier fitting and evaluation

Each of the analyses enumerated in the previous sections (e.g., A_{region}) refers to a way in which we extracted a set of interpretable time-series features from an fMRI dataset to capture different types of dynamical statistics. We sought to compare how these feature-based representations could distinguish cases from controls in each neuropsychiatric disorder using a simple measure of classification performance. We evaluated classification using linear SVMs [171], which comprise a simple classical machine learning model that can handle large input feature spaces using regularization [172]. As a robustness analysis, we subsequently evaluated classification using SVMs with an RBF kernel and random forest ensemble classifiers, finding that these two nonlinear classifiers did not substantially improve performance beyond the linear SVM kernel for A_{region}, $A_{feature}$, or A_{uni_combo} across disorders (see Fig. S1). Since the UCLA CNP cohort is characterized by class imbalances (such that there are more control participants than those with a given neuropsychiatric disorder), we applied inverse probability weighting such that samples from the minority class are assigned a larger weight when identifying the decision boundary [173]. Additionally, we evaluated classifier performance using balanced accuracy [174], which measures the average between classifier sensitivity and specificity to reflect accuracy in both the majority and minority classes. Classifiers were evaluated using stratified 10-fold cross-validation (CV) such that the hyperplane was defined based on the training data and used to predict the diagnostic labels of the unseen test data for each fold. In order to reduce outlier-driven skews in time-series feature distributions, we normalized each training fold using the scaled outlier-robust sigmoidal transformation described in Fulcher and Jones [16] and applied the same normalization to each test fold to prevent data leakage [175]. We confirmed that this reduced more distributional skew of many univariate and bivariate time-series features compared to z-score normalization as shown in Fig. S7. To ensure that results were not driven by any one specific CV split, each 10-fold CV procedure was repeated 10 times [176–178], yielding 100 out-of-sample balanced accuracy metrics per model. We report the mean $\pm$ SD tabulated across the 100 out-of-sample balanced accuracy metrics per model.

SVM classifiers were implemented in Python with the `SVC` function from *scikit-learn* [179]) with a default value of 1 for the regularization parameter, C and with the `class_weight = 'balanced'` argument for inverse probability weighting. The scaled outlier-robust sigmoidal transformation was applied using a custom Python script adapted from the Matlab-based `mixedSigmoid` normalization function from *hctsa* [15, 16]. We implemented 10 repeats of 10-fold CV using the `StratifiedKFold` function from *scikit-learn* as part of a 'pipeline', which also automatically applied normalization based on training fold parameters and evaluate balanced accuracy for each test fold. For each of the 10 repeats, we set the same random number generator state for `StratifiedKFold` such that all compared models were evaluated on the same resampled sets of test folds, which is depicted in fold-wise heatmaps in Fig. S8. We also computed the area under the receiving operator characteristic curve (AUC) for each test fold and found that this yielded comparable results to the balanced accuracy in each case, so we only report balanced accuracy in this work [173].

4.3.2 Dimensionality reduction and feature selection

While A_{region} and $A_{feature}$ both entail models with more samples than features, the other three representations—A_{uni_combo}, A_{FC}, and A_{FC_combo}—involve models with more features than samples, which risks overfitting to accumulated noise [180]. To test for such overfitting, we first evaluated two distinct methods for reducing the SVM feature space, focusing on analysis A_{uni_combo} as a starting point:

1. Principal components analysis (PCA) for dimensionality reduction; and
2. L1 (LASSO) regularization for feature selection.

First, for each case–control comparison, we applied PCA to the region $\times$ feature matrices across all cases and the corresponding control cohort using the `PCA` function from the *FactoMineR* package in R (version 2.9) [181]. We retained the first 25 principal components (PCs) to match the number of dimensions supplied to the linear SVMs for each brain region—i.e., the smallest classifier input space in our main analysis. Each case–control comparison was evaluated using these 25 PCs with the same classification procedure as described above. To complement this dimensionality reduction analysis, we also evaluated L1 (LASSO) regularization [182] which applies a penalty

term that favors shrinking the coefficients for input features to zero. The L1 optimization algorithm aims to balance between minimizing the sum of the absolute values of the feature coefficients while also minimizing the loss (i.e., still fitting the data well). We implemented this regularization using the `LinearSVC` function from 'scikit-learn' (instead of the `SVC` function used for other analyses), keeping the same regularization parameter value ($C = 1$) and class weighting ('balanced' for inverse probability weighting). The same repeated 10-fold cross-validation procedure was applied as above.

4.3.3 Statistical significance

To make statistical inferences about how a given classifier performed relative to chance level, we calculated the probability of observing a given mean balanced accuracy relative to a null distribution. For each model, we fit 1000 SVM classifiers with randomly shuffled diagnostic class labels using the same 10-repeat 10-fold CV as described above—yielding 1000 null balanced accuracy estimates per model. After confirming that all evaluated models yielded approximately normal null distributions centered at 50% (i.e., chance) balanced accuracy (Fig. S9), we derived the mean and standard deviation from each null distribution. These estimated parameters were used to compare the observed balanced accuracy with the cumulative distribution function using the `pnorm` function in R to obtain a p-value, capturing the probability of obtaining a null balanced accuracy greater than or equal to the real observed balanced accuracy. To correct for multiple comparisons within each disorder, we applied Holm–Bonferroni correction [183] to control the family-wise error rate at the $\alpha = 0.05$ level. All Holm–Bonferroni corrected p-values are denoted as p_{corr} and the number of comparisons are indicated as appropriate.

4.3.4 Volumetric analysis of group differences

In order to test whether the balanced accuracy of a given brain region was related to gray matter volume differences in case–control comparisons, we compared the number of voxels in each brain region per participant in the UCLA CNP cohort (as volumetric data is not available for the ABIDE ASD cohort). We used the *fmriprep* anatomical derivative 'aparc+aseg' parcellation and segmentation volumes that were mapped to MNI152 space using nonlinear alignment and tabulated the number of voxels in each region. Of note, these segmentation volumes correspond to the Desikan–Killiany–Tourville atlas [184] in which voxels for three brain regions (frontal pole, temporal pole, and banks of the superior temporal sulcus) were redistributed into other regions—thus precluding volumetric analyses for those three regions. For each brain region, we regressed mean volume (measured in the number of voxels) on diagnosis for each case–control comparison with an ordinary least squares model, with the resulting β coefficients listed in Supplementary Table S9. We examined the magnitude of the β coefficient estimated for each brain region to evaluate whether volumetric differences relative to controls were related to that region's balanced accuracy. The Pearson correlation coefficient was computed to measure the linear association between β coefficient magnitude and classification performance for all 82 brain regions per disorder. Additionally, to evaluate whether the overall volume of a region was related to how well its resting-state activity distinguished cases from controls, we compared the mean number of voxels per region across participants per disorder with the disorder-wise mean balanced accuracy per region using the Pearson correlation coefficient.

4.3.5 Corrected t-statistics

To quantify the change in classifier performance for pairwise FC matrices with versus without the inclusion of univariate region $\times$ feature data (from $A_{\text{uni_combo}}$), we compared the balanced accuracy distributions across all 100 test folds for the corresponding two classifier inputs per SPI. The standard t-test for group means is suboptimal in this case, as the participant overlap across test folds in the two compared models violates the assumption of independent samples [185]. We instead implemented a corrected test statistic T_{corr} designed specifically for repeated k-fold cross-validation, as defined in [186] based on the original corrected T-statistic described in Nadeau and Bengio [187]. This correction was applied using the `repkfold_ttest` function from 'correctR' package in R (version 0.1.3) [74].

4.4 Code and data availability

All data used in this study is openly accessible, with preprocessed data provided in Zenodo [188]. UCLA CNP resting-state fMRI imaging data can be downloaded from OpenNeuro with accession number ds000030 (https://openfmri.org/dataset/ds000030/). ABIDE resting-state fMRI regional time series can be downloaded from Zenodo [52]. All code used to compute univariate and pairwise time-series features, perform classification tasks, and visualize results is provided on GitHub and Zenodo [51].

5 Acknowledgements

The authors would like to thank Kieran Owens and Mac Shine for their feedback and suggestions for the manuscript and all participants and researchers involved in creating and openly sharing the fMRI datasets analyzed here. We also thank Patricia Tran and Preethom Pal for their contributions to related prior research projects.

References

1. W. B. Arthur. *The economy as an evolving complex system II*. CRC Press (2018).
2. K. G. Horton, F. A. La Sorte, D. Sheldon, et al. Phenology of nocturnal avian migration has shifted at the continental scale. *Nature climate change* **10**, 63 (2020).
3. D. S. Bassett and M. S. Gazzaniga. Understanding complexity in the human brain. *Trends in cognitive sciences* **15**, 200 (2011).
4. O. Sporns. The complex brain: Connectivity, dynamics, information. *Trends in cognitive sciences* **26**, 1066 (2022).
5. L. Pessoa. Understanding brain networks and brain organization. *Physics of life reviews* **11**, 400 (2014).
6. C. Horien, A. S. Greene, R. T. Constable, and D. Scheinost. Regions and connections: Complementary approaches to characterize brain organization and function. *The Neuroscientist* **26**, 117 (2020).
7. B. D. Fulcher. Feature-based time-series analysis. In *Feature engineering for machine learning and data analytics*, pp. 87–116. CRC press (2018).
8. S. Noble, D. Scheinost, and R. T. Constable. A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. *Neuroimage* **203**, 116157 (2019).
9. J. Hlinka, M. Paluš, M. Vejmelka, D. Mantini, and M. Corbetta. Functional connectivity in resting-state fMRI: Is linear correlation sufficient? *Neuroimage* **54**, 2218 (2011).
10. A. M. Bastos and J.-M. Schoffelen. A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Frontiers in systems neuroscience* **9**, 175 (2016).
11. Y. Zhang, M. Lucas, and F. Battiston. Higher-order interactions shape collective dynamics differently in hypergraphs and simplicial complexes. *Nature communications* **14**, 1605 (2023).
12. R. Mohanty, W. A. Sethares, V. A. Nair, and V. Prabhakaran. Rethinking measures of functional connectivity via feature extraction. *Scientific reports* **10**, 1298 (2020).
13. P. Prado, S. Moguilner, J. A. Mejía, et al. Source space connectomics of neurodegeneration: One-metric approach does not fit all. *Neurobiology of disease* p. 106047 (2023).
14. C. Gentili, I. A. Cristea, E. Ricciardi, et al. Not in one metric: Neuroticism modulates different resting state metrics within distinctive brain regions. *Behavioural brain research* **327**, 34 (2017).

15. B. D. Fulcher, M. A. Little, and N. S. Jones. Highly comparative time-series analysis: The empirical structure of time series and their methods. *Journal of The Royal Society Interface* **10**, 20130048 (2013). Number: 83 Publisher: Royal Society.

16. B. D. Fulcher and N. S. Jones. hctsa: Computational framework for automated time-series phenotyping using massive feature extraction. *Cell systems* **5**, 527 (2017). Number: 5.

17. O. M. Cliff, A. G. Bryant, J. T. Lizier, N. Tsuchiya, and B. D. Fulcher. Unifying pairwise interactions in complex dynamics. *Nature Computational Science* pp. 1–11 (2023).

18. N. H. Barbara, T. R. Bedding, B. D. Fulcher, S. J. Murphy, and T. Van Reeth. Classifying Kepler light curves for 12000 A and F stars using supervised feature-based machine learning. *Monthly notices of the Royal Astronomical Society* **514**, 2793 (2022).

19. G. Shafiei, B. D. Fulcher, B. Voytek, et al. Neurophysiological signatures of cortical micro-architecture. *Nature communications* **14** (2023).

20. J. Fallon, P. G. D. Ward, L. Parkes, et al. Timescales of spontaneous fMRI fluctuations relate to structural connectivity in the brain. *Network neuroscience* **4**, 788 (2020). Number: 3.

21. H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on signal processing* **26**, 43 (1978).

22. A. C. Linke, L. E. Mash, C. H. Fong, et al. Dynamic time warping outperforms Pearson correlation in detecting atypical functional connectivity in autism spectrum disorders. *Neuroimage* **223**, 117383 (2020).

23. R. J. Meszlényi, P. Hermann, K. Buza, V. Gál, and Z. Vidnyánszky. Resting state fMRI functional connectivity analysis using dynamic time warping. *Frontiers in neuroscience* **11**, 75 (2017).

24. A. N. Santana, I. Cifre, C. N. De Santana, and P. Montoya. Using deep learning and resting-state fMRI to classify chronic pain conditions. *Frontiers in neuroscience* **13**, 1313 (2019).

25. D. Lei, W. H. Pinaya, J. Young, et al. Integrating machining learning and multimodal neuroimaging to detect schizophrenia at the level of the individual. *Human brain mapping* **41**, 1119 (2020).

26. Z. Dai, C. Yan, Z. Wang, et al. Discriminative analysis of early Alzheimer's disease using multi-modal imaging and multi-level characterization with multi-classifier (M3). *Neuroimage* **59**, 2187 (2012).

27. X. Ding, Y. Yang, E. A. Stein, and T. J. Ross. Combining multiple resting-state fMRI features during classification: Optimized frameworks and their application to nicotine addiction. *Frontiers in human neuroscience* **11**, 362 (2017).

28. L. Xia, A. S. Malik, and A. R. Subhani. A physiological signal-based method for early mental-stress detection. In *Cyber-Enabled Intelligence*, pp. 259–289. Taylor & Francis (2019).

29. J. R. Sato, M. Q. Hoexter, A. Fujita, and L. A. Rohde. Evaluation of pattern recognition and feature extraction methods in ADHD prediction. *Frontiers in systems neuroscience* **6**, 68 (2012).

30. Z. Sha, M. Xia, Q. Lin, et al. Meta-connectomic analysis reveals commonly disrupted functional architectures in network modules and connectors across brain disorders. *Cerebral cortex* **28**, 4179 (2018).

31. B. S. Hall, R. N. Moda, and C. Liston. Glucocorticoid mechanisms of functional connectivity changes in stress-related neuropsychiatric disorders. *Neurobiology of stress* **1**, 174 (2015).

32. M. Markicevic, B. D. Fulcher, C. Lewis, et al. Cortical excitation:inhibition imbalance causes abnormal brain network dynamics as observed in neurodevelopmental disorders. *Cerebral cortex* **30**, 4922 (2020).

33. K. H. Taber, R. A. Hurley, and S. C. Yudofsky. Diagnosis and treatment of neuropsychiatric disorders. *Annual review of medicine* **61**, 121 (2010).

34. A. Surís, R. Holliday, and C. S. North. The evolution of the classification of psychiatric disorders. *Behavioral sciences* **6**, 5 (2016).

35. E. Feczko, O. Miranda-Dominguez, M. Marr, et al. The heterogeneity problem: Approaches to identify psychiatric subtypes. *Trends in cognitive sciences* **23**, 584 (2019).

36. D. A. Regier, W. E. Narrow, D. E. Clarke, et al. DSM-5 field trials in the united states and canada, part ii: Test-retest reliability of selected categorical diagnoses. *American journal of psychiatry* **170**, 59 (2013).

37. R. Freedman, D. A. Lewis, R. Michels, et al. The initial field trials of DSM-5: New blooms and old thorns. *American journal of psychiatry* **170**, 1 (2013).

38. R. Poldrack, E. Congdon, W. Triplett, et al. A phenome-wide examination of neural and cognitive function. *Scientific data* **3**, 160110 (2016).

39. A. Di Martino, C.-G. Yan, Q. Li, et al. The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry* **19**, 659 (2014).

40. A. Di Martino, D. O'Connor, B. Chen, et al. Enhancing studies of the connectome in autism using the Autism Brain Imaging Data Exchange II. *Scientific data* **4**, 170010 (2017).

41. B. W. Brunton and M. Beyeler. Data-driven models in human neuroscience and neuroengineering. *Current opinion in neurobiology* **58**, 21 (2019).

42. L. Zhang, M. Wang, M. Liu, and D. Zhang. A survey on deep learning for neuroimaging-based brain disorder analysis. *Frontiers in neuroscience* **14**, 779 (2020).

43. W. Yin, L. Li, and F.-X. Wu. Deep learning for brain disorder diagnosis based on fMRI images. *Neurocomputing* **469**, 332 (2022).

44. M. S. Breen, D. J. Stein, and D. S. Baldwin. Systematic review of blood transcriptome profiling in neuropsychiatric disorders: Guidelines for biomarker discovery. *Human psychopharmacology: Clinical and experimental* **31**, 373 (2016).

45. W. Yassin, H. Nakatani, Y. Zhu, et al. Machine-learning classification using neuroimaging data in schizophrenia, autism, ultra-high risk and first-episode psychosis. *Translational psychiatry* **10**, 278 (2020).

46. N. Khalili-Mahani, S. A. Rombouts, M. J. van Osch, et al. Biomarkers, designs, and interpretations of resting-state fMRI in translational pharmacological research: A review of state-of-the-art, challenges, and opportunities for studying brain chemistry. *Human brain mapping* **38**, 2276 (2017).

47. C.-W. Woo, L. J. Chang, M. A. Lindquist, and T. D. Wager. Building better biomarkers: Brain models in translational neuroimaging. *Nature neuroscience* **20**, 365 (2017).

48. Y. Chen. Altered functional dynamics gradient in schizophrenia with cigarette smoking. *Cerebral cortex* **33**, 7185 (2023).

49. K. Specht. Current challenges in translational and clinical fmri and future directions. *Frontiers in psychiatry* **10**, 924 (2020).

50. P. J. Uhlhaas and W. Singer. Neuronal dynamics and neuropsychiatric disorders: Toward a translational paradigm for dysfunctional large-scale networks. *Neuron* **75**, 963 (2012).

51. A. G. Bryant, K. Aquino, L. Parkes, A. Fornito, and B. D. Fulcher. DynamicsAndNeuralSystems/fMRI$_{FeaturesDisorders.https://doi.org/10.5281/zenodo.10467891(2023).}$

52. N. Traut, G. Lemaître, K. Heuer, et al. Imaging-psychiatry challenge rfMRI data (2020). URL: https://doi.org/10.5281/zenodo.3625740.

53. C. H. Lubba, S. S. Sethi, P. Knaute, et al. catch22: CAnonical Time-series CHaracteristics. *Data mining and knowledge discovery* **33**, 1821 (2019).

54. T. Henderson, A. G. Bryant, and B. D. Fulcher. Never a dull moment: Distributional properties as a baseline for time-series classification. *arXiv* (2023).

55. Y. Zhou, K. Wang, Y. Liu, et al. Spontaneous brain activity observed with functional magnetic resonance imaging as a potential biomarker in neuropsychiatric disorders. *Cognitive neurodynamics* **4**, 275 (2010).

56. Q.-H. Zou, C.-Z. Zhu, Y. Yang, et al. An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: Fractional ALFF. *Journal of neuroscience methods* **172**, 137 (2008).

57. A. Kottaram, L. A. Johnston, L. Cocchi, et al. Brain network dynamics in schizophrenia: Reduced dynamism of the default mode network. *Human brain mapping* **40**, 2212 (2019).

58. M. J. Rosa, L. Portugal, T. Hahn, et al. Sparse network-based models for patient classification using fMRI. *Neuroimage* **105**, 493 (2015).

59. X. Wang, A. Wirth, and L. Wang. Structure-based statistical features and multivariate time series clustering. In *Seventh IEEE international conference on data mining (ICDM 2007)*, pp. 351–360. IEEE (2007).

60. T. Kaufmann, K. C. Skåtun, D. Alnæs, et al. Disintegration of sensorimotor brain networks in schizophrenia. *Schizophrenia bulletin* **41**, 1326 (2015).

61. W. Xie, C.-K. Peng, C.-C. Huang, et al. Functional brain lateralization in schizophrenia based on the variability of resting-state fMRI signal. *Progress in neuro-psychopharmacology and biological psychiatry* **86**, 114 (2018).

62. A. K. Easson and A. R. McIntosh. BOLD signal variability and complexity in children and adolescents with and without autism spectrum disorder. *Developmental cognitive neuroscience* **36**, 100630 (2019).

63. J. D. Power, M. Plitt, T. O. Laumann, and A. Martin. Sources and implications of whole-brain fMRI signals in humans. *Neuroimage* **146**, 609 (2017).

64. N. K. Aurich, J. O. Alves Filho, A. M. Marques da Silva, and A. R. Franco. Evaluating the reliability of different preprocessing steps to estimate graph theoretical measures in resting state fMRI data. *Frontiers in neuroscience* **9**, 48 (2015).

65. J. Sheng, L. Zhang, J. Feng, et al. The coupling of BOLD signal variability and degree centrality underlies cognitive functions and psychiatric diseases. *Neuroimage* **237**, 118187 (2021).

66. A. El-Gazzar, M. Quaak, L. Cerliani, et al. A hybrid 3DCNN and 3DC-LSTM based model for 4D spatio-temporal fMRI data: An ABIDE autism classification study. In L. Zhou, D. Sarikaya, S. M. Kia, et al., eds., *OR 2.0 Context-Aware Operating Theaters and Machine Learning in Clinical Neuroimaging*, Lecture Notes in Computer Science, pp. 95–102. Springer International Publishing, Cham (2019).

67. J. Nielsen, B. Zielinski, P. Fletcher, et al. Multisite functional connectivity MRI classification of autism: ABIDE results. *Frontiers in human neuroscience* **7** (2013).

68. X. Yang, P. Schrader, and N. Zhang. A deep neural network study of the ABIDE repository on autism spectrum classification. *International journal of advanced computer science and applications* **11** (2020).

69. H. Jamalabadi, S. Alizadeh, M. Schönauer, C. Leibold, and S. Gais. Classification based hypothesis testing in neuroscience: Below-chance level classification rates and overlooked statistical properties of linear parametric classifiers. *Human brain mapping* **37**, 1842 (2016).

70. M. Greicius. Resting-state functional connectivity in neuropsychiatric disorders. *Current opinion in neurology* **21**, 424 (2008).

71. T. Takamura and T. Hanakawa. Clinical utility of resting-state functional connectivity magnetic resonance imaging for mood and cognitive disorders. *Journal of neural transmission* **124**, 821 (2017).

72. G. Tononi. An information integration theory of consciousness. *BMC neuroscience* **5**, 1 (2004).

73. M. Oizumi, S.-i. Amari, T. Yanagawa, N. Fujii, and N. Tsuchiya. Measuring integrated information from the decoding perspective. *PLoS computational biology* **12**, e1004654 (2016).

74. T. Henderson. hendersontrent/correctr: v0.1.2 (2022). URL: https://doi.org/10.5281/zenodo.7450146.

75. F. Petitjean, A. Ketterlin, and P. Gançarski. A global averaging method for dynamic time warping, with applications to clustering. *Pattern recognition* **44**, 678 (2011).

76. D. Schultz and B. Jain. Nonsmooth analysis and subgradient methods for averaging in dynamic time warping spaces. *Pattern recognition* **74**, 340 (2018).

77. D. S. Bassett, C. H. Xia, and T. D. Satterthwaite. Understanding the emergence of neuropsychiatric disorders with network neuroscience. *Biological psychiatry: Cognitive neuroscience and neuroimaging* **3**, 742 (2018).

78. I. Rish and G. A. Cecchi. Functional network disruptions in schizophrenia. *Biological networks and pathway analysis* pp. 479–504 (2017).

79. K. E. Stephan, K. J. Friston, and C. D. Frith. Dysconnection in schizophrenia: From abnormal synaptic plasticity to failures of self-monitoring. *Schizophrenia bulletin* **35**, 509 (2009).

80. L. Liu, Y.-P. Wang, Y. Wang, P. Zhang, and S. Xiong. An enhanced multi-modal brain graph network for classifying neuropsychiatric disorders. *Medical image analysis* **81**, 102550 (2022).

81. T. E. Mwansisya, A. Hu, Y. Li, et al. Task and resting-state fMRI studies in first-episode schizophrenia: A systematic review. *Schizophrenia research* **189**, 9 (2017).

82. Y. Zang, T. Jiang, Y. Lu, Y. He, and L. Tian. Regional homogeneity approach to fMRI data analysis. *Neuroimage* **22**, 394 (2004).

83. G. Rodriguez-Bermudez and P. J. Garcia-Laencina. Analysis of EEG signals using nonlinear dynamics and chaos: A review. *Applied mathematics & information sciences* **9**, 2309 (2015).

84. S. A. Akar, S. Kara, S. Agambayev, and V. Bilgiç. Nonlinear analysis of EEGs of patients with major depression during different emotional states. *Computers in biology and medicine* **67**, 49 (2015).

85. Y. Ma, W. Shi, C.-K. Peng, and A. C. Yang. Nonlinear dynamical analysis of sleep electroencephalography using fractal and entropy approaches. *Sleep medicine reviews* **37**, 85 (2018).

86. G. Shafiei, R. D. Markello, R. Vos de Wael, et al. Topographic gradients of intrinsic dynamics across neocortex. *eLife* **9**, e62116 (2020).

87. M. Markicevic, O. Sturman, J. Bohacek, et al. Neuromodulation of striatal D1 cells shapes BOLD fluctuations in anatomically connected thalamic and cortical regions. *eLife* **12**, e78620 (2023).

88. Y. Luo, D. Dong, H. Huang, et al. Associating multimodal neuroimaging abnormalities with the transcriptome and neurotransmitter signatures in schizophrenia. *Schizophrenia bulletin* (2023).

89. G.-Z. Wang, T. G. Belgard, D. Mao, et al. Correspondence between resting-state activity and brain gene expression. *Neuron* **88**, 659 (2015).

90. M. J. Hoptman, X.-N. Zuo, P. D. Butler, et al. Amplitude of low-frequency oscillations in schizophrenia: A resting state fMRI study. *Schizophrenia research* **117**, 13 (2010).

91. P. Wang, J. Yang, Z. Yin, et al. Amplitude of low-frequency fluctuation (ALFF) may be associated with cognitive impairment in schizophrenia: A correlation study. *BMC psychiatry* **19**, 1 (2019).

92. Y. Tang, K. Chen, Y. Zhou, et al. Neural activity changes in unaffected children of patients with schizophrenia: A resting-state fMRI study. *Schizophrenia research* **168**, 360 (2015).

93. S. M. Silverstein and B. P. Keane. Perceptual organization impairment in schizophrenia and associated brain mechanisms: Review of research from 2005 to 2010. *Schizophrenia bulletin* **37**, 690 (2011).

94. J. Gong, J. Wang, X. Luo, et al. Abnormalities of intrinsic regional brain activity in first-episode and chronic schizophrenia: a meta-analysis of resting-state functional MRI. *Journal of psychiatry and neuroscience* **45**, 55 (2020).

95. M.-J. Liang, Q. Zhou, K.-R. Yang, et al. Identify changes of brain regional homogeneity in bipolar disorder and unipolar depression using resting-state fMRI. *PLoS one* **8**, e79999 (2013).

96. M. A. Cerullo, C. M. Adler, M. P. Delbello, and S. M. Strakowski. The functional neuroanatomy of bipolar disorder. *International review of psychiatry* **21**, 314 (2009).

97. M. Giraldo-Chica and N. D. Woodward. Review of thalamocortical resting-state fMRI studies in schizophrenia. *Schizophrenia research* **180**, 58 (2017).

98. J. Hua, N. I. Blair, A. Paez, et al. Altered functional connectivity between sub-regions in the thalamus and cortex in schizophrenia patients measured by resting state bold fMRI at 7T. *Schizophrenia research* **206**, 370 (2019).

99. P. Seeman and T. Lee. Antipsychotic drugs: Direct correlation between clinical potency and presynaptic action on dopamine neurons. *Science* **188**, 1217 (1975).

100. O. D. Howes, A. J. Montgomery, M.-C. Asselin, et al. Elevated striatal dopamine function linked to prodromal signs of schizophrenia. *Archives of general psychiatry* **66**, 13 (2009).

101. F. X. Castellanos, D. S. Margulies, C. Kelly, et al. Cingulate-precuneus interactions: A new locus of dysfunction in adult attention-deficit/hyperactivity disorder. *Biological psychiatry* **63**, 332 (2008).

102. K. Wengler, A. T. Goldberg, G. Chahine, and G. Horga. Distinct hierarchical alterations of intrinsic neural timescales account for different manifestations of psychosis. *eLife* **9**, e56151 (2020).

103. L. C. Uscătescu, M. Kronbichler, S. Said-Yürekli, et al. Intrinsic neural timescales in autism spectrum disorder and schizophrenia. a replication and direct comparison study. *Schizophrenia* **9**, 18 (2023).

104. T. Watanabe, G. Rees, and N. Masuda. Atypical intrinsic neural timescale in autism. *eLife* **8**, e42256 (2019).

105. E. Nozari, M. A. Bertolero, J. Stiso, et al. Macroscopic resting-state brain dynamics are best described by linear models. *Nature biomedical engineering* pp. 1–17 (2023).

106. D. D. Garrett, N. Kovacevic, A. R. McIntosh, and C. L. Grady. Blood oxygen level-dependent signal variability is more than just noise. *Journal of neuroscience* **30**, 4914 (2010).

107. D. D. Garrett, U. Lindenberger, R. D. Hoge, and C. J. Gauthier. Age differences in brain signal variability are robust to multiple vascular controls. *Scientific reports* **7**, 1 (2017).

108. V. Scarapicchia, E. L. Mazerolle, J. D. Fisk, L. J. Ritchie, and J. R. Gawryluk. Resting state BOLD variability in Alzheimer's disease: A marker of cognitive decline or cerebrovascular status? *Frontiers in aging neuroscience* **10**, 39 (2018).

109. D. D. Garrett, S. M. Epp, A. Perry, and U. Lindenberger. Local temporal variability reflects functional integration in the human brain. *Neuroimage* **183**, 776 (2018).

110. G. Baracchini, Y. Zhou, J. da Silva Castanheira, et al. The biological role of local and global fMRI BOLD signal variability in human brain organization. *bioRxiv* (2023).

111. M. Wehrheim, J. Faskowitz, J. C. G. Alanis, A.-L. Schubert, and C. Fiebach. Reliability of variability and complexity measures for task and task-free BOLD fMRI (2023).

112. J. Arnedo, D. M. Svrakic, C. Del Val, et al. Uncovering the hidden risk architecture of the schizophrenias: Confirmation in three independent genome-wide association studies. *American journal of psychiatry* **172**, 139 (2015).

113. M. Bleich-Cohen, S. Jamshy, H. Sharon, et al. Machine learning fMRI classifier delineates subgroups of schizophrenia patients. *Schizophrenia research* **160**, 196 (2014). Number: 1-3.

114. L. Niu, K. Fang, S. Han, C. Xu, and X. Sun. Resolving heterogeneity in schizophrenia, bipolar i disorder, and attention-deficit/hyperactivity disorder through individualized structural covariance network analysis. *Cerebral Cortex* p. bhad391 (2023).

115. S. S. Sethi, V. Zerbi, N. Wenderoth, A. Fornito, and B. D. Fulcher. Structural connectome topology relates to regional BOLD signal dynamics in the mouse brain. *Chaos: An interdisciplinary journal of nonlinear science* **27**, 047405 (2017). Number: 4 Publisher: American Institute of Physics.

116. J. Massey. Causality, feedback and directed information. In *Proceedings of the International Symposium on Information Theory and Its Applications (ISITA-90)*, pp. 303–305 (1990).

117. G. Kramer. *PhD Dissertation: Directed information for channels with feedback*. Citeseer (1998).

118. B. He and Z. Liu. Multimodal functional neuroimaging: Integrating functional MRI and EEG/MEG. *IEEE reviews in biomedical engineering* **1**, 23 (2008).

119. J. F. Enwright III and D. A. Lewis. Similarities in cortical transcriptome alterations between schizophrenia and bipolar disorder are related to the presence of psychosis. *Schizophrenia bulletin* **47**, 1442 (2021).

120. A. Arnatkevičiūtė, A. Fornito, J. Tong, et al. Linking GWAS to pharmacological treatments for psychiatric disorders. *medRxiv* pp. 2022–05 (2022).

121. Y. Liu, M. S. Mellem, H. Gonzalez, et al. Highly predictive transdiagnostic features shared across schizophrenia, bipolar disorder, and adhd identified using a machine learning based approach. *bioRxiv* p. 453951 (2018).

122. A. Vabalas, E. Gowen, E. Poliakoff, and A. J. Casson. Machine learning algorithm validation with a limited sample size. *PLoS one* **14**, e0224365 (2019).

123. A. A. Pulini, W. T. Kerr, S. K. Loo, and A. Lenartowicz. Classification accuracy of neuroimaging biomarkers in attention-deficit/hyperactivity disorder: Effects of sample size and circular analysis. *Biological psychiatry: Cognitive neuroscience and neuroimaging* **4**, 108 (2019).

124. M. A. Just, V. L. Cherkassky, A. Buchweitz, T. A. Keller, and T. M. Mitchell. Identifying autism from neural representations of social interactions: Neurocognitive markers of autism. *PLoS one* **9**, e113879 (2014). Publisher: Public Library of Science.

125. X. Yang, M. S. Islam, and A. M. A. Khaled. Functional connectivity magnetic resonance imaging classification of autism spectrum disorder using the multisite ABIDE dataset. In *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pp. 1–4 (2019). ISSN: 2641-3604.

126. N. Traut, K. Heuer, G. Lemaître, et al. Insights from an autism imaging biomarker challenge: Promises and threats to biomarker discovery. *Neuroimage* **255**, 119171 (2022).

127. M. Ingalhalikar, S. Shinde, A. Karmarkar, et al. Functional connectivity-based prediction of autism on site harmonized ABIDE dataset. *IEEE transactions on biomedical engineering* **68**, 3628 (2021). Conference Name: IEEE Transactions on Biomedical Engineering.

128. O. Artiles, Z. Al Masry, and F. Saeed. Confounding effects on the performance of machine learning analysis of static functional connectivity computed from rs-fMRI multi-site data. *Neuroinformatics* pp. 1–18 (2023).

129. Q. Li, W. Zhao, L. Palaniyappan, and S. Guo. Atypical hemispheric lateralization of brain function and structure in autism: a comprehensive meta-analysis study. *Psychological medicine* pp. 1–12 (2023).

130. B. Wan, S.-J. Hong, R. A. Bethlehem, et al. Diverging asymmetry of intrinsic functional organization in autism. *bioRxiv* pp. 2023–04 (2023).

131. A. M. Buch, P. E. Vértes, J. Seidlitz, et al. Molecular and network-level mechanisms explaining individual differences in autism spectrum disorder. *Nature neuroscience* **26**, 650 (2023).

132. M. A. Reiter, A. Jahedi, A. J. Fredo, et al. Performance of machine learning classification models of autism using resting-state fMRI is contingent on sample heterogeneity. *Neural computing and applications* **33**, 3299 (2021).

133. R. Ramasubbu, M. R. Brown, F. Cortese, et al. Accuracy of automated classification of major depressive disorder as a function of symptom severity. *Neuroimage: Clinical* **12**, 320 (2016).

134. I. Malkiel, G. Rosenman, L. Wolf, and T. Hendler. Self-supervised transformers for fMRI representation. In *International Conference on Medical Imaging with Deep Learning*, pp. 895–913. PMLR (2022).

135. K. L. Peterson, R. Sanchez-Romero, R. D. Mill, and M. W. Cole. Regularized partial correlation provides reliable functional connectivity estimates while correcting for widespread confounding. *bioRxiv* pp. 2023–09 (2023).

136. M. Venkatesh, J. Jaja, and L. Pessoa. Comparing functional connectivity matrices: A geometry-aware approach applied to participant identification. *Neuroimage* **207**, 116398 (2020).

137. I. M. Johnstone and D. M. Titterington. Statistical challenges of high-dimensional data. *Philosophical transactions of the Royal Society A: Mathematical, physical and engineering sciences* **367**, 4237 (2009).

138. G. Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE transactions on information theory* **14**, 55 (1968).

139. G. V. Trunk. A problem of dimensionality: A simple example. *IEEE transactions on pattern analysis and machine intelligence* pp. 306–307 (1979).

140. B. T. Yeo, F. M. Krienen, J. Sepulcre, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology* (2011).

141. B. Wang, A. M. Mezlini, F. Demir, et al. Similarity network fusion for aggregating data types on a genomic scale. *Nature methods* **11**, 333 (2014).

142. J. Y. Hansen, G. Shafiei, K. Voigt, et al. Multimodal, multiscale connectivity blueprints of the cerebral cortex. *bioRxiv* pp. 2022–12 (2022).

143. N. W. Bailey, B. D. Fulcher, B. Caldwell, et al. Uncovering a stability signature of brain dynamics associated with meditation experience using massive time-series feature extraction. *Neural networks* (2023).

144. A. H. Williams, T. H. Kim, F. Wang, et al. Unsupervised discovery of demixed, low-dimensional neural dynamics across multiple timescales through tensor component analysis. *Neuron* **98**, 1099 (2018).

145. E. J. Müller, B. R. Munn, K. M. Aquino, J. M. Shine, and P. A. Robinson. The music of the hemispheres: Cortical eigenmodes as a physical basis for large-scale brain activity and connectivity patterns. *Frontiers in human neuroscience* **16**, 1062487 (2022).

146. J. C. Pang, K. M. Aquino, M. Oldehinkel, et al. Geometric constraints on human brain function. *Nature* pp. 1–9 (2023).

147. X. Zhu, H. Li, H. T. Shen, et al. Fusing functional connectivity with network nodal information for sparse network pattern learning of functional brain networks. *Information fusion* **75**, 131 (2021).

148. J. Y. Hansen, G. Shafiei, K. Voigt, et al. Integrating multimodal and multiscale connectivity blueprints of the human cerebral cortex in health and disease. *PLoS biology* **21**, e3002314 (2023).

149. R. L. Peach, A. Arnaudon, J. A. Schmidt, et al. HCGA: Highly comparative graph analysis for network phenotyping. *Patterns* **2**, 100227 (2021). Number: 4.

150. Y. Xu, X. Long, J. Feng, and P. Gong. Interacting spiral wave patterns underlie complex brain dynamics and are related to cognitive processing. *Nature human behaviour* pp. 1–20 (2023).

151. R. G. Townsend and P. Gong. Detection and analysis of spatiotemporal patterns in brain activity. *PLoS computational biology* **14**, e1006643 (2018).

152. A. Santoro, F. Battiston, M. Lucas, G. Petri, and E. Amico. Higher-order connectomics of human brain function reveals local topological signatures of task decoding, individual identification, and behavior. *bioRxiv* pp. 2023–12 (2023).

153. A. Bagnall, J. Lines, A. Bostrom, J. Large, and E. Keogh. The great time series classification bake off: A review and experimental evaluation of recent algorithmic advances. *Data mining and knowledge discovery* **31**, 606 (2017).

154. O. Esteban, C. J. Markiewicz, R. W. Blair, et al. fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods* **16**, 111 (2019). Number: 1 Publisher: Nature Publishing Group.

155. R. H. R. Pruim, M. Mennes, D. van Rooij, et al. ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. *Neuroimage* **112**, 267 (2015).

156. K. M. Aquino, B. D. Fulcher, L. Parkes, K. Sabaroedin, and A. Fornito. Identifying and removing widespread signal deflections from fMRI data: Rethinking the global signal regression problem. *Neuroimage* **212**, 116614 (2020).

157. L. Parkes, B. Fulcher, M. Yücel, and A. Fornito. An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional MRI. *Neuroimage* **171**, 415 (2018).

158. K. M. Aquino, Fulcher, Ben, Oldham, Stuart, et al. On the intersection between data quality and dynamical modelling of large-scale fMRI signals. *Neuroimage* **256**, 119051 (2022).

159. R. S. Desikan, F. Ségonne, B. Fischl, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968 (2006).

160. B. Fischl, D. H. Salat, E. Busa, et al. Whole brain segmentation. *Neuron* **33**, 341 (2002).

161. M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith. FSL. *Neuroimage* **62**, 782 (2012).

162. J. D. Power, K. A. Barnes, A. Z. Snyder, B. L. Schlaggar, and S. E. Petersen. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* **59**, 2142 (2012).

163. X. Liu, H. Yang, B. Becker, et al. Disentangling age-and disease-related alterations in schizophrenia brain network using structural equation modeling: A graph theoretical study based on minimum spanning tree. *Human brain mapping* **42**, 3023 (2021).

164. W. Huijbers, K. R. Van Dijk, M. M. Boenniger, R. Stirnberg, and M. M. Breteler. Less head motion during MRI under task than resting-state conditions. *Neuroimage* **147**, 111 (2017).

165. E. P. K. Pua, S. Barton, K. Williams, J. M. Craig, and M. L. Seal. Individualised MRI training for paediatric neuroimaging: A child-focused approach. *Developmental cognitive neuroscience* **41**, 100750 (2020).

166. X.-N. Zuo, A. Di Martino, C. Kelly, et al. The oscillating brain: Complex and reliable. *Neuroimage* **49**, 1432 (2010).

167. J. A. Turner, E. Damaraju, T. G. Van Erp, et al. A multi-site resting state fMRI study on the amplitude of low frequency fluctuations in schizophrenia. *Frontiers in neuroscience* **7**, 137 (2013).

168. A. M. Golestani, J. B. Kwinta, Y. B. Khatamian, and J. J. Chen. The effect of low-frequency physiological correction on the reproducibility and specificity of resting-state fMRI metrics: Functional connectivity, ALFF, and ReHo. *Frontiers in neuroscience* **11**, 546 (2017).

169. T. Henderson and B. D. Fulcher. Feature-based time-series analysis in R using the theft package (2022).

170. O. Cliff, A. G. Bryant, B. Fulcher, and A. Nguyen. DynamicsAndNeuralSystems/pyspi (2023). URL: https://doi.org/10.5281/zenodo.8223340.

171. C. Cortes and V. Vapnik. Support-vector networks. *Machine learning* **20**, 273 (1995).

172. B. Schölkopf and A. J. Smola. *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. MIT press (2002).

173. P. Thölke, Y.-J. Mantilla-Ramos, H. Abdelhedi, et al. Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data. *Neuroimage* **277**, 120253 (2023).

174. K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann. The balanced accuracy and its posterior distribution. In *2010 20th International Conference on Pattern Recognition*, pp. 3121–3124. IEEE (2010).

175. N. Bussola, A. Marcolini, V. Maggio, G. Jurman, and C. Furlanello. AI slipping on tiles: Data leakage in digital pathology. In *Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10–15, 2021, Proceedings, Part I*, pp. 167–182. Springer (2021).

176. G. Varoquaux, P. R. Raamana, D. A. Engemann, et al. Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *Neuroimage* **145**, 166 (2017).

177. J. Li, R. Kong, R. Liégeois, et al. Global signal regression strengthens association between resting-state functional connectivity and behavior. *Neuroimage* **196**, 126 (2019).

178. B. Larsen, Z. Cui, A. Adebimpe, et al. A developmental reduction of the excitation:inhibition ratio in association cortex during adolescence. *Science advances* **8** (2022).

179. F. Pedregosa, G. Varoquaux, A. Gramfort, et al. Scikit-learn: Machine learning in Python. *Journal of machine learning research* **12**, 2825 (2011).

180. M. Verleysen and D. François. The curse of dimensionality in data mining and time series prediction. In *Computational Intelligence and Bioinspired Systems: 8th International Work-Conference on Artificial Neural Networks, IWANN 2005, Vilanova i la Geltrú, Barcelona, Spain, June 8-10, 2005. Proceedings 8*, pp. 758–770. Springer (2005).

181. S. Lê, J. Josse, and F. Husson. FactoMineR: an R package for multivariate analysis. *Journal of statistical software* **25**, 1 (2008).

182. R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **67**, 91 (2005).

183. S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics* pp. 65–70 (1979).

184. A. Klein and J. Tourville. 101 labeled brain images and a consistent human cortical labeling protocol. *Frontiers in neuroscience* **6**, 171 (2012).

185. M. P. Fay and M. A. Proschan. Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. *Statistics surveys* **4**, 1 (2010).

186. R. R. Bouckaert and E. Frank. Evaluating the replicability of significance tests for comparing learning algorithms. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 3–12. Springer (2004).

187. C. Nadeau and Y. Bengio. Inference for the generalization error machine learning. *Machine learning* **52**, 239 (2003).

188. A. G. Bryant, K. Aquino, L. Parkes, A. Fornito, and B. D. Fulcher. All data to support this manuscript. https://zenodo.org/doi/10.5281/zenodo.10431854 (2023).

Supplementary Figures

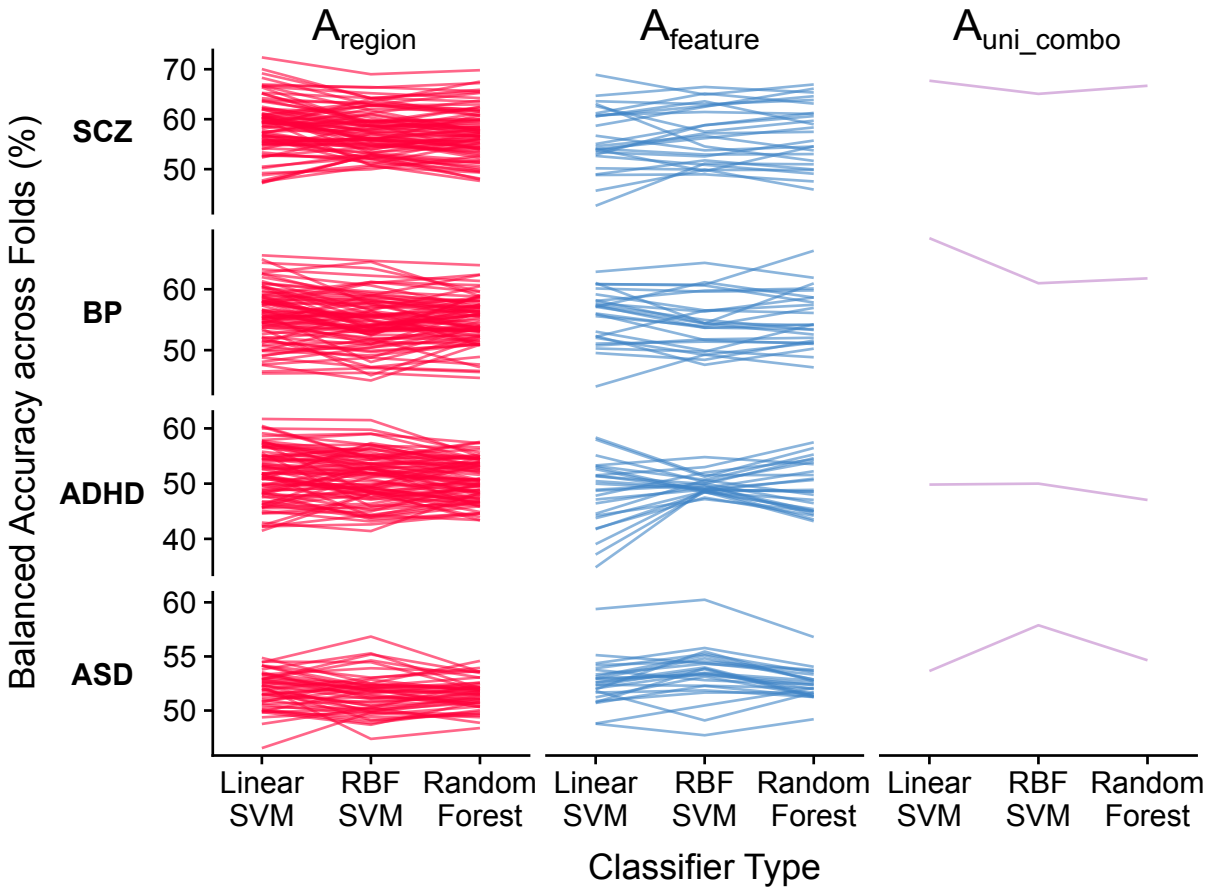

Figure S1. Comparing linear SVM with RBF SVM and random forest ensembles supports the use of a linear classifier. Here, we compare balanced accuracy for the three analyses based on intra-regional properties—A_{region}, $A_{feature}$, and A_{uni_combo}—using linear SVM, RBF SVM, and random forest classifiers. Lines indicate the same model compared across the three classifier types; for example, the left pericalcarine will map to one line in the A_{region} column to aid visualization.

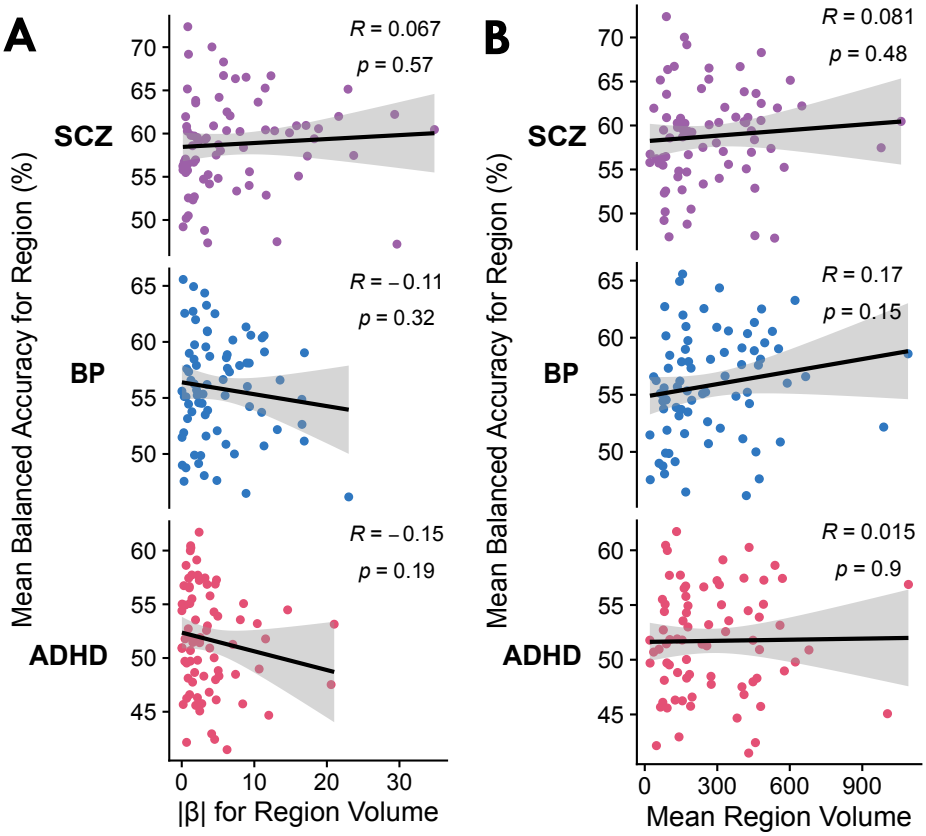

Figure S2. Region-wise classification balanced accuracy is not associated with volumetric differences across clinical groups. A For each brain region in the UCLA CNP dataset, the mean balanced accuracy is plotted relative to the absolute β coefficient estimated from ordinary least squares regression of region volume on diagnosis per clinical group. Pearson correlation estimates (R) and corresponding p-values are annotated in the top right corners. **B** As in A, for each brain region, the mean balanced accuracy is plotted relative to the average region volume (measured in number of voxels) across all participants in the UCLA CNP cohort. Pearson correlation estimates, R, and corresponding p-values are shown in the top right corners.

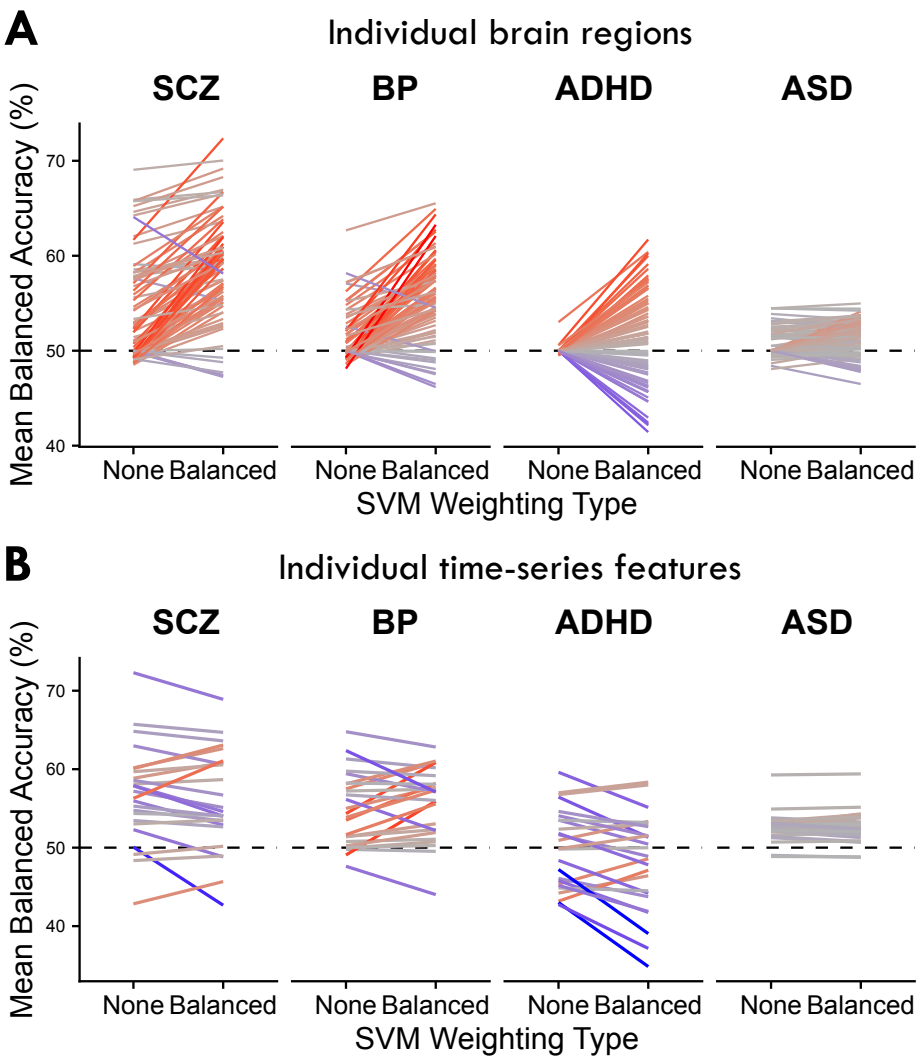

Figure S3. Comparing classification performance with versus without inverse probability weighting.
A For each disorder, the mean balanced accuracy per brain region is shown with no weighting ('None') or inverse probability weighting ('Balanced'). Lines correspond to each of 82 brain regions for SCZ, BP, and ADHD and each of 48 brain regions for ASD. Colors are included as a visual aid to highlight the difference in performance between the two weighting types, with red corresponding to higher balanced accuracy with inverse probability weighting and blue corresponding to lower balanced accuracy with inverse probability weighting. **B** For each disorder, the mean balanced accuracy per univariate time-series feature is shown with no weighting ('None') or inverse probability weighting ('Balanced'), as in A.

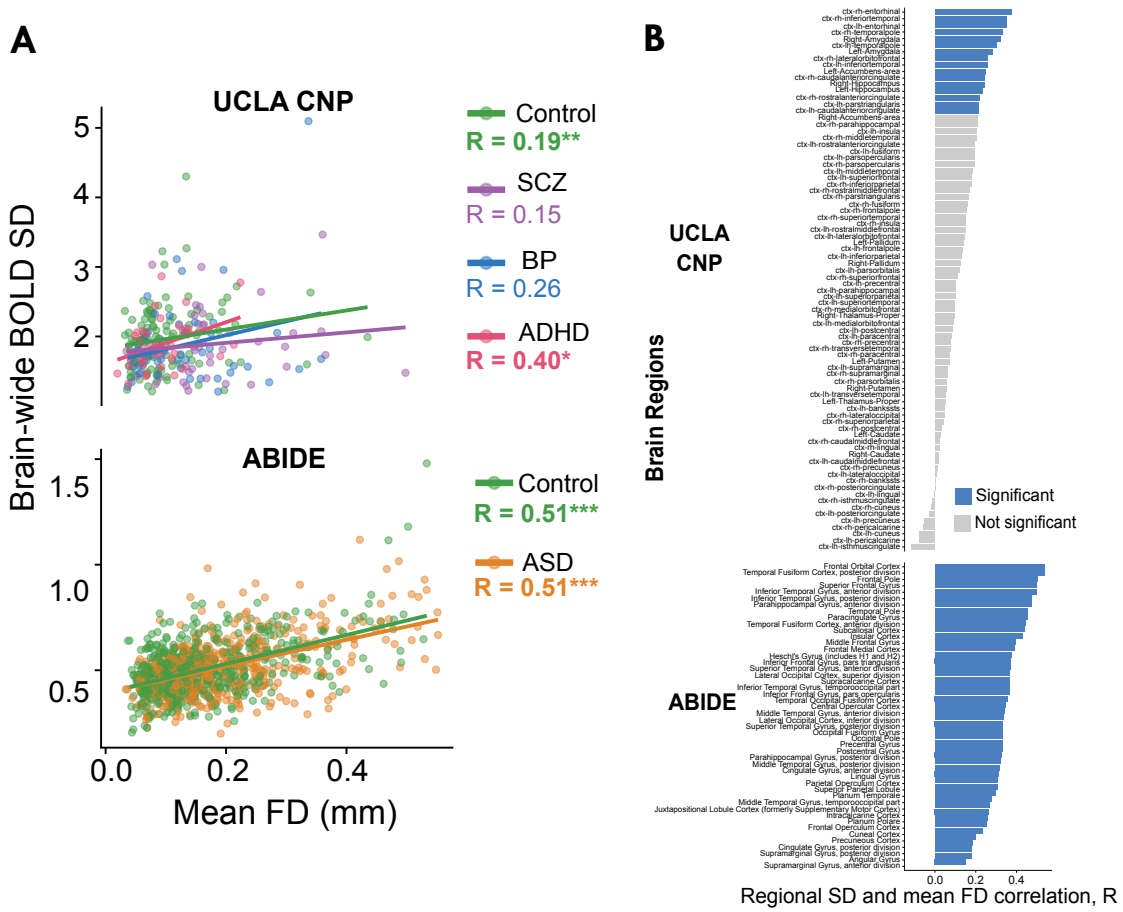

Figure S4. Head movement in the scanner is associated with whole-brain and regional BOLD SD in some disorders. A. The BOLD standard deviation (SD) showed the strongest positive correlation with mean FD of any feature examined, with $R = 0.51$ for both controls ($p = 3.3 \times 10^{-39}$) and ASD participants ($p = 5.0 \times 10^{-35}$) in the ABIDE cohort. The brain-wide average BOLD SD is plotted against the mean FD across participants in the UCLA CNP cohort (upper) and the ABIDE cohort (lower), with the Pearson correlation estimates R shown to the right in each plot. * indicates $p < 0.05$, ** indicates $p < 0.01$, and *** indicates $p < 0.001$. **B.** In order to compare BOLD SD with mean framewise displacement (FD), for each brain region, we calculated the Pearson correlation coefficient (R) between BOLD SD and mean FD—partitioned by study (UCLA CNP versus ABIDE) to account for data preprocessing differences. The Pearson correlation estimates, R, are shown for all brain regions per study, ordered by correlation magnitude and colored to indicate statistical significance (blue indicating $p_{\text{corr}} < 0.05$, corrected across 82 brain regions for UCLA CNP and 48 brain regions for ABIDE). All 48 brain regions exhibited a significant correlation in the ABIDE dataset, and 16 out of 82 brain regions exhibited a significant correlation in the UCLA CNP dataset.

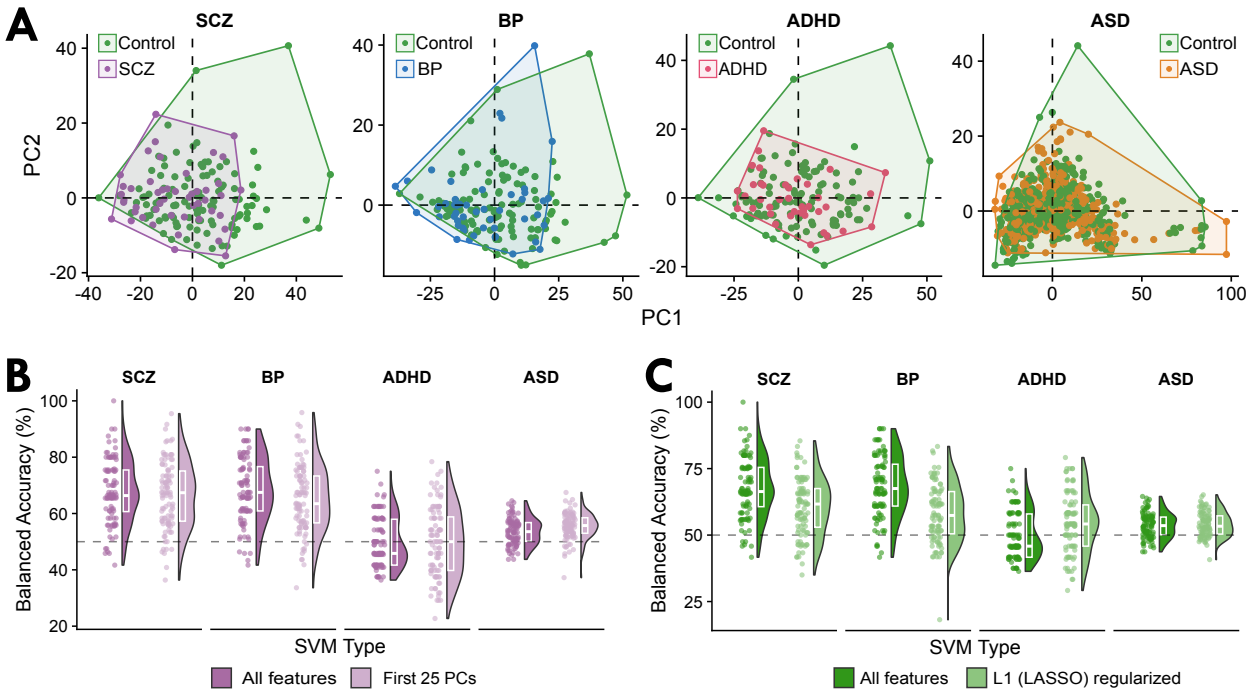

Figure S5. Linear dimensionality reduction and regularization approaches did not improve out-of-sample classification for the univariate region $\times$ feature classifiers, $\mathbf{A}_{\mathbf{uni_combo}}$. **A** For each disorder, individual scores for the first two PCs are plotted, with points colored according to diagnosis. Shaded areas reflect convex hulls encapsulating all points for each diagnostic group. Note that each PCA was computed separately for each case–control comparison, so PC1 and PC2 scores are not directly comparable across clinical groups. **B** For each case–control comparison, we compare the out-of-sample balanced accuracy across the 100 repeats $\times$ folds using all region $\times$ feature variables (left, dark purple) versus using only scores for the first 25 PCs (right, light purple). Points are randomly jittered along the horizontal axis to aid visualization. Box plots are included in each half-violin to indicate the median and interquartile range of the distribution. **C** For each case–control comparison, we compare the out-of-sample balanced accuracy across the 100 repeats$\times$folds using default regularization (left, dark green) versus L1 regularization (right, light green).

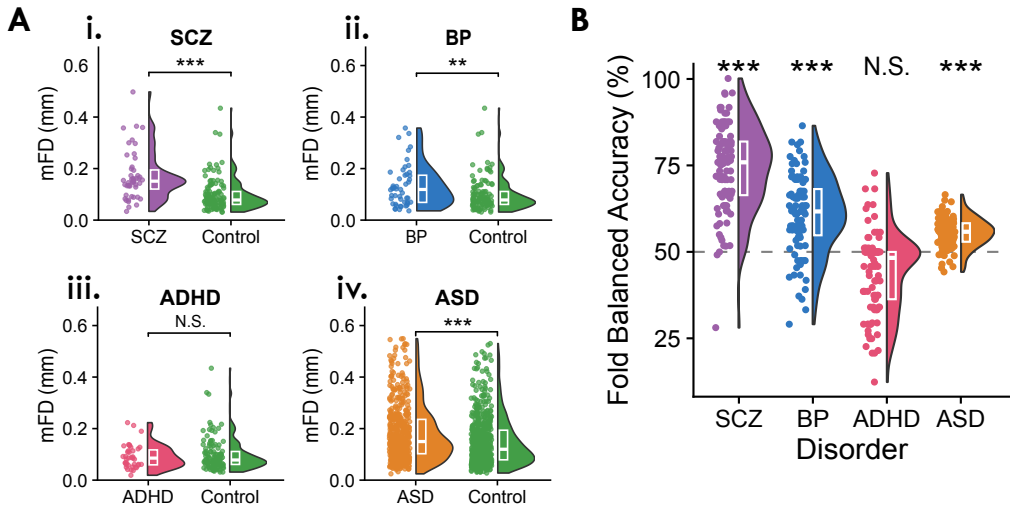

Figure S6. The median mFD values are higher for SCZ, BP, and ASD patients than controls, and mFD values can distinguish cases from controls. **A.** In order to evaluate the relevance of head movement to clinical diagnosis, we compared the mean framewise displacement (mFD) distributions across participants for each disorder relative to controls using violin plots. Differences in group mFD distributions were assessed with the Wilcoxon rank sum test, with the resulting significance level indicated above the violin plots. ** indicates $p < 0.01$, *** indicates $p < 0.001$, and N.S. indicates not significant. **B.** For each disorder, we tested case–control classification performance using just head motion on its own, with the same cross-validated linear SVM procedure applied throughout all analyses in this study. The balanced accuracies are plotted across all 100 folds $\times$ repeats for each disorder as dots, with the distributions summarized as violin plots, inside of which boxplots show the median and interquartile range. The mean balanced accuracy per group was compared against the same pooled empirical null distribution from the set of 25 univariate time-series features to yield a p-value, with the significance level indicated above each violin plot.

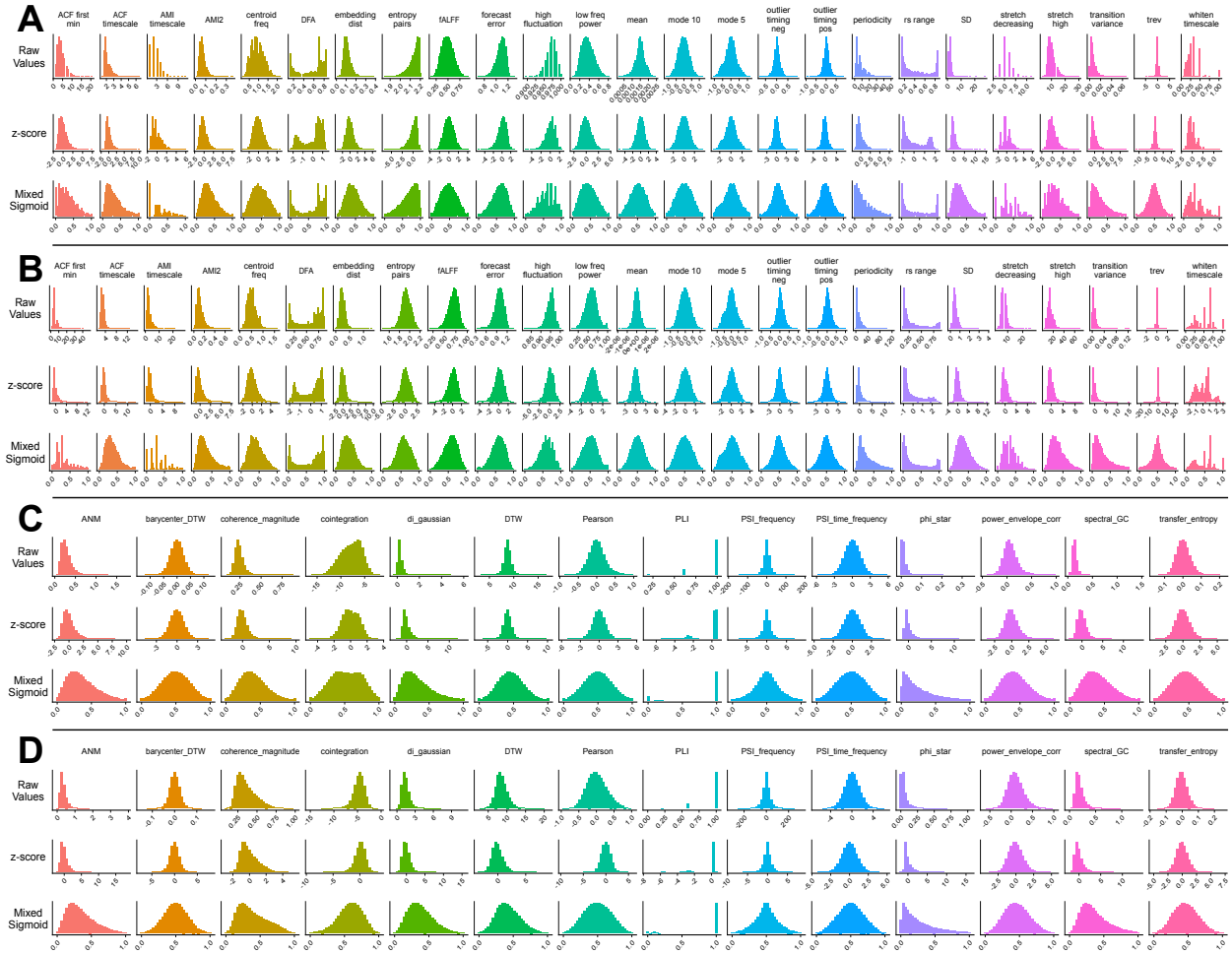

Figure S7. Comparing normalization methods supports the use of the outlier-robust mixed sigmoid method. Whole-brain univariate time-series feature value distributions are shown for all participants in the UCLA CNP cohort (**A**) or ABIDE cohort (**B**) with no normalization (upper row), z-score normalization (middle row), and outlier-robust mixed sigmoid normalization (bottom row; see Methods Sec. 4.3.1 for description). Whole-brain pairwise SPI value distributions are shown for all participants in the UCLA CNP cohort (**C**) or ABIDE cohort (**D**) with no normalization (upper row), z-score normalization (middle row), and outlier-robust mixed sigmoid normalization (bottom row).

A

Study	Disorder	% control participants overall	Mean % controls across all folds/repeats	Min % controls across all folds/repeats	Max % controls across all folds/repeats
UCLA CNP	SCZ	71.1	71.1	68.8	75
	BP	70.7	70.7	68.8	75
	ADHD	75.2	75.2	73.3	80
ABIDE	ASD	52.3	52.3	52.2	53

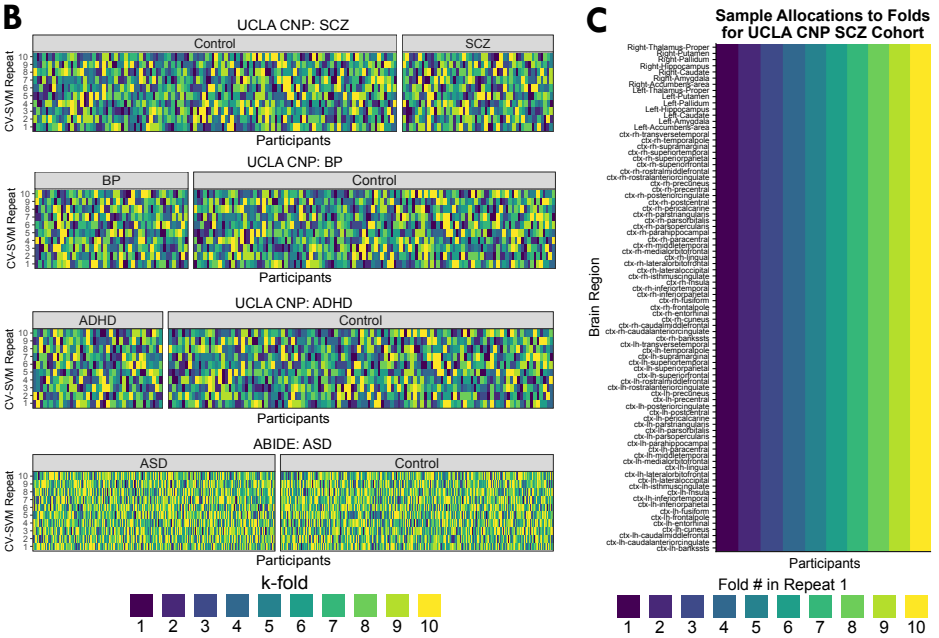

Figure S8. Participants are allocated into the same set of folds and repeats across models to enable fair comparison of performance across models. A Here we denote the proportion of control participants relative to each disorder in the overall dataset as well as the mean percent of controls across all stratified test folds, confirming that the proportion is preserved across cross-validation folds. We also include the minimum and maximum proportion of controls across all test folds. **B** Here we show the test fold assignment for each participant across repeats for one example model (the resting-state activity in the left caudal anterior cingulate, selected at random for demonstration). Colors indicate the test fold number assigned to each participant across the ten repeats per disorder, confirming that participants were re-shuffled into different test folds across resamples. **C** To confirm that participants were allocated into the same set of test folds across all models for fair comparison, we visualize fold assignments for all brain regions in Repeat #1 for SCZ–control comparisons. Participants are ordered along the horizontal axis by fold assignment to underscore that Fold #1 in the left banks of the superior temporal sulcus ('ctx-lh-bankssts', bottom row) is identical to that of all brain regions through the right thalamus ('Right-Thalamus-Proper', top row).

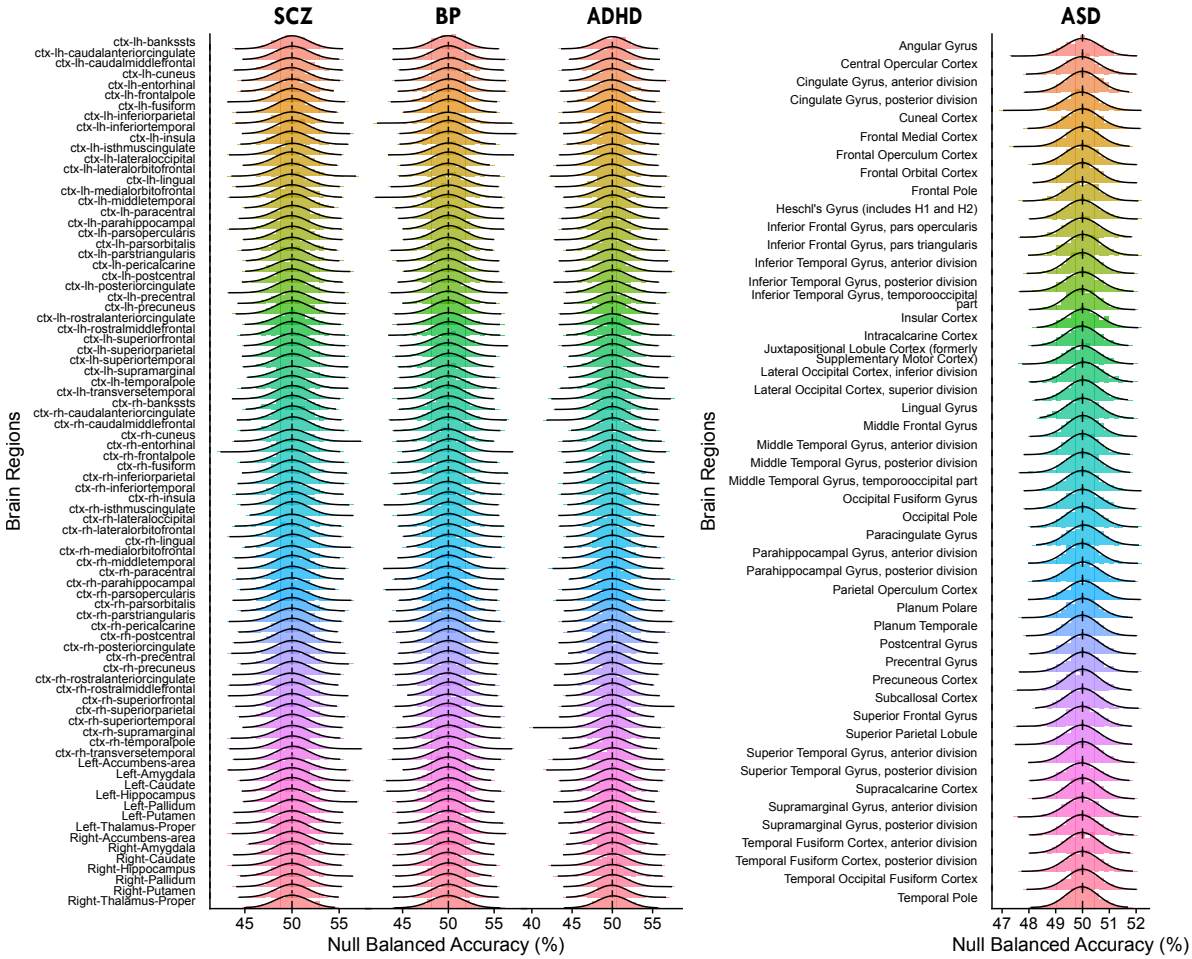

Figure S9. Null balanced accuracy metrics are approximately normally distributed across brain regions in all clinical cohorts. The distribution of 1000 null balanced accuracy values (cross-validated over 10 repeats of 10 folds each) is shown for each brain region by disorder. Empirical null values are visualized as histograms, with the probability density overlaid as black lines for the normal distribution based on the mean and standard deviation for each null distribution. These normal distributions were used to compute p-values for each brain region; the same procedure was also applied for univariate and pairwise time-series features. ASD brain regions are plotted separately from the SCZ, BP, and ADHD brain regions as different parcellation atlases were used in the two cohort datasets (ABIDE versus UCLA CNP). The dashed vertical line marks 50% balanced accuracy in all plots.