Random Matrix Theory Tools for the Predictive Analysis of Functional Magnetic Resonance Imaging Examinations

Derek Berger[a], **Gurpreet S. Matharoo**[b,c], **Jacob Levman**[a*,d,e,f]

[a]St. Francis Xavier University, Department of Computer Science, 4130 University Avenue, Antigonish, Nova Scotia, Canada, B2G 2W5

[b]St. Francis Xavier University, ACENET, 4130 University Avenue, Antigonish, Nova Scotia, Canada, B2G 2W5

[c]St. Francis Xavier University, Department of Physics, 4130 University Avenue, Antigonish, Nova Scotia, Canada, B2G 2W5

[d]Athinoula A. Martinos Center for Biomedical Imaging, 149 Thirteenth Street, Suite 2301, Charlestown, Massachusetts, United States, 02129

[e]Harvard Medical School, Department of Radiology, 25 Shattuck Street Boston, Massachusetts, United States, 02115

[f]Nova Scotia Health Authority, Research Affiliate, Nova Scotia, Canada

Abstract.

Purpose

Random matrix theory (RMT) is an increasingly useful tool for understanding large, complex systems. Prior studies have examined functional Magnetic Resonance Imaging (fMRI) scans using tools from RMT, with some success. However, RMT computations are highly sensitive to a number of analytic choices, and the robustness of findings involving RMT remain in question. We systematically investigate the usefulness of RMT on a wide variety of fMRI datasets using a rigorous predictive framework.

Approach

We develop open-source software to efficiently compute RMT features from fMRI images, and examine the cross-validated predictive potential of eigenvalue and RMT-based features ("eigenfeatures") with classic machine-learning classifiers. We systematically vary pre-processing extent, normalization procedures, RMT unfolding procedures, and feature selection, and compare the impact of these analytic choices on the distributions of cross-validated prediction performance for each combination of dataset binary classification task, classifier, and feature. To deal with class imbalance, we use the area under the receiver operating characteristic curve (AUROC) as the main performance metric.

Results

Across all classification tasks and analytic choices, we find RMT- and eigenvalue-based "eigenfeatures" to have predictive utility more often than not (82.4% of median AUROCs > 0.5; median AUROC range across classification tasks 0.47 - 0.64). Simple baseline reductions on source timeseries, by contrast, were less useful (58.8% of median AUROCs > 0.5, median AUROC range across classification tasks 0.42 - 0.62). Additionally, eigenfeature AUROC distributions were overall more right-tailed than baseline features, suggesting greater predictive potential. However, performance distributions were wide, and often significantly affected by analytic choices.

Conclusions

Eigenfeatures clearly have potential for understanding fMRI functional connectivity in a wide variety of scenarios. The utility of these features is strongly dependent on analytic decisions, suggesting caution when interpreting past and future studies applying RMT to fMRI. However, this study demonstrates that the inclusion of RMT statistics in fMRI investigations could improve prediction performances across a wide variety of phenomena.

Keywords: random matrix, spectral rigidity, level number variance, fMRI, classification, machine-learning.

***Jacob Levman**, jlevman@stfx.ca (primary), jlevman@mgh.harvard.edu

1 Introduction

At its most rudimentary, RMT describes the expected behavior of the eigenvalues—also often called levels—of a number of classes of random matrices.[1–3] A random matrix is, simplifying somewhat, a matrix in which the entries are drawn from independently and identically distributed (*iid*) distributions. Though first developed to describe the fluctuation of nuclei energy levels in quantum physics,[3,4] RMT has been shown to have extremely broad potential. In small-scale physical systems, RMT universalities have been observed in quantum chaotic systems, complex nuclei, atoms, molecules and disordered mesoscopic systems,[3–9] and at larger scales, RMT has been applied to atmospheric physics,[10] stock cross-correlations,[11] social networks,[12] random networks,[13] network-formation in liquids[14,15] and amorphous clusters.[16–18] Within biological systems, RMT has also been used to successfully model aspects of amino acid functional relationships,[19] synchrony in epileptic seizures,[20] and in protein-protein interactions both in different species[21] and breast cancer.[22] RMT has also been used to guide statistical decisions in principal components analyses,[23–25] and, more recently, has provided insights into the behaviors of deep neural networks.[26,27]

RMT likely has the most real-world explanatory potential when dealing with a complex system of many (hundreds or more[3]) interacting components. If such a system has a matrix representation, and the eigenvalues of this representation have a distribution similar to one predicted by RMT, it suggests the system is either highly random or chaotic. By contrast, if the observed spectra deviate significantly from RMT-predicted spectra, this suggests otherwise. A number of studies have used RMT to make such interpretations and comparisons between systems.[10,12,13,21,22,28–31]

1.1 RMT and Neurobiological Signals

In the human brain, each neuron, collection of neurons, or region of interest (ROI) is a potentially-interacting component in a complex system. RMT may have potential in describing the totality of these interactions, provided that measurements of functioning can be obtained with sufficient spatial and temporal resolution to speak to some neurobiological or neuropsychological phenomenon of interest.

For an imaging modality like fMRI, where changes in the blood-oxygenation-level-dependent (BOLD) signals are related to neural activity, RMT may be an ideal starting point, as each voxel time course (or collection of such time sources, i.e. ROIs) can be considered an interacting component of the system. Likewise, in functional connectivity analyses, statistical relationships of the BOLD ROI time-courses are investigated in the hope of gaining insights into brain function (see 32, 33 for reviews, but also 34 and 35 for challenges facing functional connectivity analyses). In this framework, each connection or correlation can be considered a system component, and the eigenvalues of such a correlation matrix can be examined from the perspective of RMT.

The earliest study taking this approach demonstrated that spectra of the correlations between electroencephalographic (EEG) signals closely resemble those of the Gaussian Orthogonal Ensemble (GOE).[28] In fMRI, RMT has been used to evaluate the quality of whole brain features extracted from fMRI data,[36,37] and in diffusion MRI to aid in the selection of the number of components to employ in principal-component reduction analysis and denoising.[24,25,37]

RMT has also been used in ROI-based fMRI functional connectivity studies to investigate differences between rest and task states,[29] between subjects with and without attention-deficit hyperactive disorder (ADHD),[30] between pain and non-pain states,[31] and between left-sided vs. right-

sided motor imagery.[38] While no differences were found in the latter study, across the first three studies, the spectra of resting or low-attention states exhibited properties close to the GOE. These findings suggest that certain aspects of psychological processes might be characterized, in part, by features computed from the eigenvalues of fMRI correlation matrices, and that these features might vary in an interpretable manner across psychological processes. If this is the case, RMT could aid in understanding the functioning of the human brain.

1.2 Eigenvalue Features

The basic insight of RMT is thus that eigenvalues alone may provide interesting information about highly complex systems. However, real systems are usually noisy, and involve a mixture of random and non-random components and interactions. RMT is statistical in nature, describing only the expected behavior of the spectra of *iid* random matrices. To this end, a number of summary statistics or "spectral observables"[3,4] can be computed from empirically observed spectra, with these spectral observables sometimes being better suited for further analysis, or for the comparison of empirical observations to theory. Two such summary statistics that have been popular[10,12,13,21,22,28–31] are the *spectral rigidity* and *level number variance* ("rigidity" and "level variance" for short; details in Section 3).

However, from a predictive standpoint, these summary statistics may eliminate predictively-useful information. If the basic RMT insight is that the eigenvalues alone can provide understanding of a system, then those eigenvalues (or other simple transformations of them) ought also to be predictively useful. We examine a number of such features in this study, and refer to both RMT-derived features like the rigidity and level variance, and non-RMT-derived features collectively as *eigenfeatures* in this study.

1.3 Functional Connectivity Reduction

The functional connectivity—often, the matrix of correlations of various collections of voxels of an fMRI scan—is *a priori* a useful representation of the fMRI data. However, the full correlation matrix between all voxels is itself often computationally infeasible to work with, and thus is reduced in various ways prior to being used in analyses.

Typical reductions include the use of a "seed" voxel or collection of voxels (see e.g. 39, 40): the mean signal over that ROI is correlated with all other N ROIs of interest to generate a reduced functional connectivity matrix (or image) with one correlation value at each non-seed ROI. This sort of reduction reduces the functional connectivity to N values, but necessarily biases the analysis to the seed ROI.

Likewise, one can work with ROI mean signal reductions, or independent component analysis reductions,[39,40] and work with the $N \times N$ matrix of the correlations of these reductions. This analysis is, *a priori*, sensitive to the choice of ROIs, and, in the case of anatomical atlases / parcellations for ROIs, may also lack theoretical justification.

RMT suggests a potentially useful reduction in the form of the eigenvalues of the voxelwise functional connectivity. An $N \times t$ matrix of N time series of length t, and where $t \ll N$ will have a symmetric correlation (or covariance) matrix with $t - 1$ non-zero, positive real-valued eigenvalues. These eigenvalues can be computed highly efficiently via transposition of the voxelwise correlation or covariance matrix.

1.4 Limitations of Previous Work

Previous studies applying RMT to functional connectivity data[30,31,38,41] took largely descriptive / explanatory approaches, noting only differences in RMT across subgroups and/or conditions. For

example Wang et al. (2016)[30] and Gu et al. (2020)[38] use a Kolmogorov-Smirnov test, and Wang et al. (2015)[41] uses a t-test to provide evidence of differences in various RMT metrics. However, statistically significant differences in a metric do not necessarily imply practical significance or predictive utility,[42] nor do they imply generalization or replicability.[43] Cross-validation, by contrast, attempts to more directly assess these properties.[44]

In addition, the previous papers do not provide publicly-accessible code to reproduce results. The extraction of the spectral rigidity and level variance is computationally-demanding and mathematically and algorithmically non-trivial, and additionally requires an *unfolding* procedure.[3,4] Unfolding is an exponential fitting procedure that involves a number of highly subjective decisions regarding outliers and the flexibility of the fitting function. These decisions are, unfortunately, often poorly or even entirely undocumented, but are known to often dramatically impact RMT conclusions.[45–49]

In addition to the flexibility in the implementation and application of RMT to empirical data, there is also flexibility introduced by analytic choices made in the complex preprocessing pipelines of fMRI.[50] These "researcher degrees of freedom",[51] in combination with the absence of reproducible code and data availability, and predominance of descriptive and explanatory approaches, raise doubts about the basic robustness and practical utility of RMT-based functional connectivity analyses.

What is missing is a rigorous, systematic, reproducible investigation of the predictive value of RMT metrics across a wide variety of data, analytic choices, RMT features, and preprocessing decisions, with RMT features being compared to simpler alternative baseline predictors. The current study aims to remedy this.

2 Datasets

2.1 Overview

We selected fMRI data publicly available on the OpenNeuro platform.[52] Selection criteria was somewhat subjective, but we required each dataset to (1) comprise 10 or more human subjects, (2) have all fMRI scans have the same spatial and temporal resolutions, (3) have fMRI images that can be split into various classes for a number of binary classification tasks, and (4), allow reasonably classifying each scan in its entirety (i.e. neither task nor event timings are needed to determine class membership, nor do any scan volumes need to be removed to make the scan classifiable). This yielded seven datasets total.

Table 1 Included fMRI dataset details. Name = Identifier for paper. Dimensions listed as M x N x P, indicate P axial slices each with dimensions M x N. TR = Time of Repetition (seconds). n_scans = Total number of images in dataset.

Name	Dimensions	Voxel size (mm)	TR	Volumes	n_scans
Aging	$74 \times 74 \times 32$	$3.0 \times 3.0 \times 4.0$	2.0	300	62
Bilingual	$100 \times 100 \times 72$	$1.8 \times 1.8 \times 1.8$	0.88	823	90
Depress	$112 \times 112 \times 25$	$2.0 \times 2.0 \times 5.0$	2.5	100	72
Learn	$64 \times 64 \times 36$	$3.0 \times 3.0 \times 3.0$	2.0	195	432
Osteo	$64 \times 64 \times 36$	$3.4 \times 3.4 \times 3.0$	2.5	300	74
Park	$80 \times 80 \times 43$	$3.0 \times 3.0 \times 3.0$	2.4	149	552
Attention	$128 \times 128 \times 70$	$1.5 \times 1.5 \times 1.5$	3.0	300	90

In order not to inundate readers with dataset details, and because this is an exploratory investigation of RMT which is not committed to any specific theory[1], we refer the reader to the original publications and/or data releases when greater detail is needed, and highlight only the most basic aspects of each dataset here. Dataset scan parameters and acquisition details are summarized in Table 1, and sample and subgroup sizes are summarized in Table 2.

[1] In fact, datasets and, later, the classification tasks, were chosen without investigation into any findings or publications associated with the data, both in order to reduce bias in our decisions, and because the original authors analytic / statistical approaches and intentions have very limited relevance to our whole-brain, voxelwise, multiverse predictive approach.

Table 2 Sizes and other details for classification task subgroups. ID = Identifier for paper. Subgroup = name of subgroup used in classification task. Subjects = Number of subjects. Task = fMRI task. ANT = Attention Network Task[53]

ID	Subgroup	Subjects	Task	Scans per Subject
Aging	older	28	resting-state	1
	younger	34	resting-state	1
Bilingual	bilingual	59	resting-state	1
	monolingual	33	resting-state	1
Depress	depression	51	resting-state	1
	control	21	resting-state	1
Learn	task	24	learn image sketches	16
	rest	24	resting-state	2
Osteo	duloxetine	19	resting-state	1
	pain	37	resting-state	1
	nopain	20	resting-state	1
Park	Parkinson's	25	ANT	12
	control	21	ANT	12
Attention	vigilant	11	resting-state	2
	nonvigilant	11	resting-state	2
	trait-attentive	11	resting-state	2
	trait-non-attentive	11	resting-state	2
	task-attentive	11	resting-state	2
	task-nonattentive	11	resting-state	2

2.2 Datasets and Classification Tasks

The *Aging* dataset[54] included rs-fMRI data for 34 subjects with a mean age of 22 years (range: 18-32 years), and 28 subjects with a mean age of 70 years (range: 61-80 years). For our analysis, we make the binary classification task the prediction of subject age-group membership, i.e. `younger v older`.

The *Bilinguality* dataset[55,56] examined English and Spanish-speaking monolinguals and multilinguals during a prolonged resting state. The original study grouped participants into three subgroups: early vs late bilinguals vs monolingual controls.[56] For our analysis, we make the binary

classification task `monolingual v bilingual`.

The *Depression* dataset[57,58] includes rs-fMRI scans from non-depressed controls and mildly- or moderately-depressed subjects. In our analysis, the binary classification task is `depress v control`.

The *Learning* dataset[59,60] has both rs-fMRI and task fMRI scans available for all subjects, and `task v rest` is chosen as the binary classification task for our analysis. The task is complex and difficult to summarize here adequately, but involved multiple phases where subjects were presented with various related abstract images, and then later tested for their memory of certain aspects of the learned images. Thus, the classification task might also be considered learning vs. rest.

The *Osteo* dataset[61] includes whole-brain rs-fMRI scans of healthy, pain-free controls (`nopain` condition), and individuals with knee osteoarthritis. Osteoarthritic patients were treated for two weeks with either placebo (`pain` condition) or duloxetine (`duloxetine` condition). We use the three binary classification tasks `nopain v duloxetine`, `nopain v pain`, and `pain v duloxetine` for our analysis.

The *Park* dataset[62] includes subjects with non-demented Parkinson's Disease (PD) and healthy controls performed a number of repetitions of the Attention Network Task (ANT[53]) during scans. The classification task for this dataset is Parkinson's versus controls, i.e. `park v ctrl`.

The *Attention* dataset[63] is a high-resolution rs-fMRI dataset including a battery of psychological measures for each subject. Since previous studies employing RMT sometimes interpreted their findings with respect to attentional processes,[30,31] we divided subjects into various high vs. low attention binary classification tasks based on median splits of subsets of the metrics available in this study.

The `vigilant v nonvigilant` task was formed based on self-report questionnaire items involving "vigilance".[64] The `trait_attend v trait_nonattend` task was formed using PANAS-X[65] items related to self-reported wakefulness and attention over the past weeks. Finally, the `task_attend v task_nonattend` classification task was constructed using the scores of the Conjunctive Continuous Performance Task (CCPT[66]). This is a behavioral performance task which requires the subject to quickly and selectively respond to only certain visual stimuli. Details allowing exact reproduction of these splits are available in this paper's source code.

3 Methods

3.1 fMRI Preprocessing

We limited preprocessing to a simple pipeline of, in order: (1) brain extraction, (2) slice timing correction, (3), motion correction, and (4), registration to the MNI152[67,68] 2mm isotropic, asymmetric template version C made available through TemplateFlow.[69]

Brain extraction was performed first with BET,[70] and then with ANTs[71] to clean up residue left behind by BET. BET results were visually inspected for each image in each dataset, and BET fractional anisotropy values were modified to ensure brain extraction was acceptable. Motion-correction was performed with FSL's MCFLIRT,[72] and slice time correction with FSL's slicetimer[73] tool. Finally, functional images were registered directly to the MNI152 template via ANTs.[71]

We limited preprocessing methods to these steps because data required for other typical preprocessing steps (e.g. field intensity maps, physiological measurements) was missing, but also because all pipeline intermediates were saved to later compare the effect of increasing degrees

of preprocessing on feature predictive utility (see Section 3.6), and it was important to limit the number of preprocessing steps to prevent excessive computational costs and comparisons.

3.2 Feature Extraction

For each preprocessing degree and fMRI image, features were extracted from all N non-constant brain voxels, where N may differ for each image. After this process, each scan yields an $N \times t$ matrix $\mathbf{M}$, where t is the number of volumes acquired. All features extracted summarize this matrix $\mathbf{M}$.

3.2.1 Raw Eigenvalues

The time and space complexities of computing the full $N \times N$ $\mathbf{M}$ matrix of Pearson correlation coefficients (and subsequent eigenvalues) are too large to be tractable. However, since transposition does not change the eigenvalues, we can use the transpose to efficiently compute these eigenvalues (see Appendix A). These raw eigenvalues (feature `eigs` in Table 3), most directly test the idea that eigenvalues alone are useful functional connectivity features.

However, not all eigenvalues may have predictive utility. We also examine as features the central eigenvalues (defined as the middle 10%, 20%, or 40% of the sorted eigenvalues, `eigsmiddle` in Table 3), and tail eigenvalues (defined as the first and last 5%, 10%, and 20% of each tail of the spectrum, `eigsminmax` in Table 3).

3.2.2 RMT Features

To compare the observed eigenvalue distribution to some RMT theoretical predictions, the eigenvalues must be unfolded.[3,4] The details and motivation behind the unfolding procedure are well documented elsewhere.[4] Practically, the unfolding process can be viewed as a smoothing and

Table 3 Feature groupings for summarization. eigs = non-RMT eigenfeatures and their combinations with other non-RMT eigenfeatures. rmt = RMT eigenfeatures and their combinations. tseries = baseline timeseries reductions.

Coarse Grouping	Fine Grouping	Feature id
eigs	eigs	`eigs`
	eigs max	`eigsminmax10`
		`eigsminmax20`
		`eigsminmax5`
	eigs middle	`eigsmiddle10`
		`eigsmiddle20`
		`eigsmiddle40`
	eigs smooth	`eigs + eigs_smooth`
		`eigs + savgol`
		`eigs_savgol`
		`eigs_smooth`
rmt	rmt + eigs	`eigs + levelvar`
		`eigs + rigidity`
		`eigs + rigidity + levelvar`
		`eigs + unfolded`
		`eigs + unfolded + levelvar`
		`eigs + unfolded + rigidity`
	rmt only	`levelvar`
		`rigidity`
		`rigidity + levelvar`
		`unfolded`
		`unfolded + levelvar`
		`unfolded + rigidity`
		`unfolded + rigidity + levelvar`
tseries	location	`T-max`
		`T-mean`
		`T-med`
		`T-min`
		`T-p05`
		`T-p95`
	scale	`T-iqr`
		`T-rng`
		`T-rrng`
		`T-std`

rescaling procedure where the originally observed spectrum $\lambda_1 \leq \cdots \leq \lambda_n$ is smoothly mapped to an unfolded spectrum $e_1, \ldots, e_n$, such that $\sum d_i/n \approx 1$, if $d_i = e_{i+1} - e_i$. For most empirical data, the true smoothing function is not known, and so must be approximated.[3,4] Typically, a polynomial

is chosen.[45]

Given empirically observed eigenvalues Λ, the spectral rigidity, $\Delta_3(L)$, is calculated for any positive real value $L < \max(\Lambda)$ as:

$$\Delta_3(L) = \left\langle \min_{A,B} \frac{1}{L} \int_c^{c+L} (\eta(\lambda) - A\lambda - B)^2 \right\rangle_c \tag{1}$$

where $\eta(\lambda)$ is the number of unfolded eigenvalues less than or equal to λ, $\langle\cdot\rangle_c$ denotes the average with respect to all starting points c, and where A and B denote the slope and intercept, respectively, of the least squares fit of a straight line to $\eta(\lambda)$ on $[c, c+L]$.[4] Viewing the unfolded eigenvalues as a timeseries, the spectral rigidity for a value L is the *average nonlinearity* of all intervals of length L over the series.

The level number variance, $\Sigma^2(L)$, or *level variance*, for short, is closely related to the spectral rigidity,[3] and is calculated as:

$$\Sigma^2(L) = \left\langle \eta^2(L, c) \right\rangle_c - \langle \eta(L, c) \rangle_c^2 \tag{2}$$

where $\eta(L, c)$ is the number of unfolded eigenvalues in $[c, c+L]$, and where c, L, and $\langle\cdot\rangle_c$ are as above.[4] Viewing the unfolded eigenvalues as an irregular timeseries, the level number variance is the variation of the number of samples in all intervals of length L over the series.

As eigenfeatures, we compute the unfolded eigenvalues, and the spectral rigidity and level number variance for all $L \in \{1, 2, \ldots 20\}$. These are `unfolded`, `rigidity`, and `levelvar`, respectively, in Table 3.

Trimming Variants. As the unfolding procedure operates on the sorted eigenvalues, and in-

volves fitting a smooth polynomial or exponential to these values, extreme values are often omitted from fitting.[48,49] The number of eigenvalues to trim is subjective, and unfortunately, with a large number of different source matrices, it is too destructive to naively use the same hard criterion (e.g. largest 10%, largest three) for all unfoldings.

We develop and test four trimming variants: no trimming, precision-based, largest, and middle, with less subjective criteria for determining trimming thresholds. The details of these trimming procedures are straightforward, and can be found in Appendix B, or the source code.

Unfolding Variants. Following any trimming, the eigenvalues are fit with a polynomial. Because the choice of degree has been known to dramatically impact certain analyses involving the spectral rigidity or level variance,[45–49] we compute the rigidity and level variance with all possible combinations of our trimming procedures, and unfolding polynomial degrees of 3, 5, 7, and 9.

Software Release. Because the computation of these statistics is non-trivial, and to make the unfolding procedure more transparent and reproducible, we developed and release a separate, open-source Python library `empyricalRMT`.[74] The library allows for efficient, parallel computation of the spectral rigidity and level variance via Monte Carlo methods, and automatically ensures convergence of the statistics to user-specified tolerances. The library also makes available various other functions useful for empirical RMT analyses, such as unfolding and trimming functions, and plotting facilities for classic RMT ensembles.

3.2.3 Smoothed Eigenfeatures

When computing RMT eigenfeatures, the polynomial fitting during unfolding has a smoothing effect, as do the additional averaging $\langle\ \rangle_c$ operations. However, it is possible other, simpler transformations of the eigenvalues (e.g. uniform or Savitsky-Golay smoothing) might have equal or

greater predictive utility. We thus also test smoothed variants of the eigenvalues as predictive features: the sorted eigenvalues are smoothed with either a uniform (moving average) filter or Savitsky-Golay, using window sizes of 3, 5, 7, and 9 to yield 8 total features (`eigs_smooth` and `eigs_savgol` in Table 3).

3.2.4 Feature Combinations

A combination of RMT and non-RMT eigenfeatures could be more predictive than either alone. We test this with a variety of eigenfeature combinations, with a focus on combinations that involve the simplest features (e.g. raw eigenvalues, unfolded eigenvalues) and then features involving additional processing (e.g. smoothed eigenvalues, level variance, rigidity). Combined features are formed by concatenation, so that if we have features $f_1, \ldots, f_n$ with dimensions $p_1, \ldots, p_n$, then the combined feature is $[f_1; \ldots; f_n]$ with $\sum p_i$ dimensions. The final combinations chosen can be found in Table 3, where the "+" symbol indicates concatenation.

3.2.5 Slicing Variants

The largest, middle, or smallest eigenvalues could be most useful in characterizing any given system. E.g. if the smallest eigenvalues correspond to random / noise aspects of a system, but differences in the nature of the system noise most differentiate between systems, then the smallest eigenvalues may have the most predictive utility. Likewise, the L value in each RMT eigenfeature defines the degree of locality at which we summarize the spectrum: with small L-values, the spectral rigidity summarizes the non-linearity of eigenvalues that are relatively close to each other in magnitude. At large values of L, the rigidity summarizes the long-range non-linearity of the spectrum.

Since predictive utility may vary with summary locality, or with the region of the original spectrum, we investigate various front, middle, and end contiguous slices of each eigenfeature in all analyses, where the size of each slice is either the first or last 5%, 10%, or 20% of the full eigenfeatures for non-middle slices, or the middle 10%, 20%, or 40% of the full eigenfeature for the middle eigenvalues.

For combined features, slicing variants are also computed, with slicing performed first on each feature to be combined. E.g. if we have features $f_1, \ldots, f_n$ with dimensions $p_1, \ldots, p_n$, then the sliced max 25% feature for each component feature, $\hat{f}_i$ is the last $p_i/4$ elements of f_i, and the combined sliced feature is $[\hat{f}_1;\ \ldots;\ \hat{f}_n]$.

3.2.6 Baseline Features

If RMT- or eigenvalue-based features fail to predictively outperform features that are simple summary statistics of the fMRI data, it is difficult to justify the greater computational and interpretational complexities of the former. We compute ten simple statistical reductions of the voxel time series ("tseries" features in Table 3) to help assess the relative value of RMT features.

The baseline reductions used as statistical summaries were: robust measures of location (mean, max, min); non-robust measures of location (median, 95^{th} percentile, 5^{th} percentile); non-robust measures of scale (standard deviation and the range—`T-rng` in Table 3); and robust measures of scale (interquartile range and difference between 95^{th} percentile and 5^{th} percentile—`T-rrng` in Table 3). Each statistic reduces the fMRI data along the voxel dimension, yielding a feature of t dimensions (e.g. the "mean" feature, `T-mean`, is the usual global mean signal).

We do not take slice variants of these baseline features, since the baseline features are still time series. However, we do also evaluate a number of smoothing degrees of these features, to account

for noise and to be somewhat similar to the various smoothing variants for the eigenfeatures. Each baseline feature is tested with a degree of uniform smoothing, where the size of the smoothing window is either 1 (no smoothing), 2, 4, 8, or 16.

3.3 Classifiers

We use a variety of standard machine-learning classifiers available in the Scikit-learn[75] Python library to solve each classification task. We use a gradient-boosted decision tree (GBDT), random forest (RF) classifier, support vector classifier (SVC) with radial basis function, and k-nearest neighbors classifiers with k equal to 3, 5, and 9 (KNN3, KNN5, and KNN9, respectively), in all cases with the default hyperparameter values. We originally also attempted to test a simple logistic regression classifier, but found that this model frequently failed to converge for a number of features, even after increasing iterations significantly.

3.4 Preprocessing Levels

Each preprocessing intermediate in the (1) brain extraction, (2) slice time correction, (3) motion correction, and (4) template registration pipeline was saved and used for entirely separate feature analyses. E.g. all features were extracted from an fMRI image prepared with one of four levels or degrees of preprocessing, where preprocessing level k includes preprocessing steps 1 to k, inclusive, and starting at $k = 1$.

3.5 Normalization

Because of the exponential distribution of the eigenfeatures, normalization is somewhat non-trivial, and a simple standardization or min-max normalization is ineffective. We instead first apply a logarithm to all eigenfeatures, and then test each classifier with un-normalized or min-max normalized

versions of the log-features. Baseline features are also tested with un-normalized and min-max normalized versions, but without any log transform.

3.6 Multiverse Analysis

We perform a *multiverse analysis*[76] to assess the overall predictive potential of the various eigenfeatures across all previously-mentioned analytic choices. That is, for each combination of comparison task, classifier, and analytic choices, we use 5-fold cross validation, and use the mean AUROC (mAUROC) across folds as our metric to evaluate feature predictive utility.

The AUROC metric was chosen because it handles class imbalances present in the various datasets, and is naturally normalized and interpretable such that mAUROC < 0.5 indicates predictive performance worse than guessing, and mAUROC > 0.5 indicates positive predictive utility.[77] However, we also collected as performance metrics the accuracy, adjusted accuracy (proportion of samples in the entire dataset largest class minus accuracy), and the F1 score. A complete table with these additional metrics is available with the data for this study.

Each non-baseline feature is thus evaluated, for each dataset classification task and each classifier, exactly 1280 times: there are 4 preprocessing levels $\times$ 2 normalization methods $\times$ 4 trimming choices $\times$ 4 unfolding degrees $\times$ 10 slice choices. Each baseline feature is evaluated exactly 40 times for each comparison task and classifier: 4 preprocessing levels $\times$ 2 normalization methods $\times$ 5 smoothing window sizes. In total, with all the datasets and features examined in this study, this yields 2,053,920 mAUROC values to summarize, with 5 main *analytic factors*— preprocessing, normalization, trimming, unfolding/smoothing degree, and slicing—to consider.

4 Results

4.1 Overview

Summarizing this quantity of data requires some caution. Measures of location or scale, even when robust, are, for the most part, misleading and uninformative when distributions are skewed. In most cases we find skewed mAUROC distributions, and so have chosen primarily to present our findings visually, with kernel density estimates. Because aggregating across datasets with different class imbalances can be extremely misleading, even with a metric like the mAUROC,[78] we restrict such summaries to Table 4 only, and note that the table is intended as a *supplement only* to the distribution plots: no conclusions should be drawn about feature performance from the table values alone.

In addition, a large number of interactions are possible between each analytic factor in this study. For example, it could be that trimming impacts the overall mAUROC distribution only at a certain preprocessing level and for certain feature slicing. However, as there are too many potential interactions to present, we limit our presentation to main effects.

To aid in summarizing the performances of the large number of different features, we also summarize patterns of results across two abstract groupings of similar features ("coarse", and "fine") shown in Table 3. Thus, for example, figures depicting the coarse feature grouping "eigs" in fact depict all observed mAUROC values for the fine feature groupings of raw, max, middle, and smoothed eigenvalues, and the fine feature grouping "eigs smooth", for example, depicts all observed mAUROC values for both the `eigs_savgol` and `eigs_smooth` features.

Table 4 Numerical summaries of feature mAUROCs across predictable comparisons, and all combinations of analytic choices, sorted by 95% percentile (robust max) value. Bold values indicate column "best" values, when reasonable.

Feature	mean	min	5%	50%	95%	max	std
eigs + eigs_smooth	**0.583**	0.188	0.408	0.567	**0.788**	0.906	0.111
eigs + savgol	0.580	0.188	0.408	**0.563**	0.785	0.913	0.110
unfolded	0.568	0.150	0.400	0.555	0.775	0.931	0.108
unfolded + levelvar	0.569	0.150	0.400	0.555	0.775	0.925	0.108
unfolded + rigidity	0.568	0.150	0.401	0.555	0.775	0.925	0.108
unfolded + rigidity + levelvar	0.568	0.150	0.401	0.555	0.775	0.931	0.108
eigs + rigidity + levelvar	0.564	0.202	0.401	0.548	0.763	**0.937**	0.105
eigs + unfolded	0.564	0.202	0.401	0.549	0.763	0.919	0.105
eigs + unfolded + levelvar	0.564	0.202	0.400	0.548	0.763	0.933	0.105
eigs + unfolded + rigidity	0.564	0.202	0.401	0.548	0.763	0.920	0.105
eigs + levelvar	0.563	0.202	0.400	0.548	0.758	0.914	0.104
eigs	0.563	0.202	0.401	0.548	0.756	0.875	0.104
eigs + rigidity	0.563	0.202	0.400	0.548	0.756	0.882	0.104
eigs_smooth	0.569	0.188	0.417	0.556	0.755	0.907	0.102
T-p05	0.526	0.249	0.340	0.500	0.754	0.856	0.110
eigs_savgol	0.566	0.188	0.410	0.553	0.748	0.909	0.102
eigsminmax20	0.556	0.245	0.411	0.544	0.738	0.906	0.098
eigsminmax10	0.550	0.231	0.409	0.538	0.737	0.906	0.095
eigsminmax5	0.549	0.213	0.400	0.539	0.729	0.888	0.096
T-mean	0.557	0.208	**0.427**	0.558	0.717	0.814	0.085
levelvar	0.534	0.075	0.367	0.527	0.712	0.879	0.101
eigsmiddle40	0.547	**0.324**	0.419	0.533	0.697	0.785	0.084
rigidity + levelvar	0.529	0.154	0.377	0.523	0.696	0.888	0.094
rigidity	0.529	0.154	0.378	0.523	0.694	0.888	0.094
T-rrng	0.545	0.281	0.384	0.542	0.688	0.759	0.090
eigsmiddle20	0.545	0.321	0.419	0.533	0.686	0.798	0.080
T-iqr	0.494	0.175	0.312	0.490	0.682	0.759	0.099
eigsmiddle10	0.541	0.323	0.411	0.530	0.681	0.755	0.081
T-std	0.513	0.252	0.392	0.500	0.679	0.729	0.086
T-rng	0.538	0.279	0.392	0.539	0.678	0.779	0.092
T-max	0.540	0.275	0.385	0.546	0.678	0.754	0.095
T-med	0.539	0.221	0.399	0.541	0.665	0.702	0.077
T-p95	0.524	0.294	0.402	0.514	0.653	0.693	0.077
T-min	0.504	0.321	0.426	0.500	0.582	0.698	0.046

4.2 Classification Tasks

4.2.1 Non-Predictable Comparisons

It is important, to reduce figure clutter and complexity, to exclude a classification task or an analytic factor from summaries if there is no evidence that the classification task is solvable in general, or

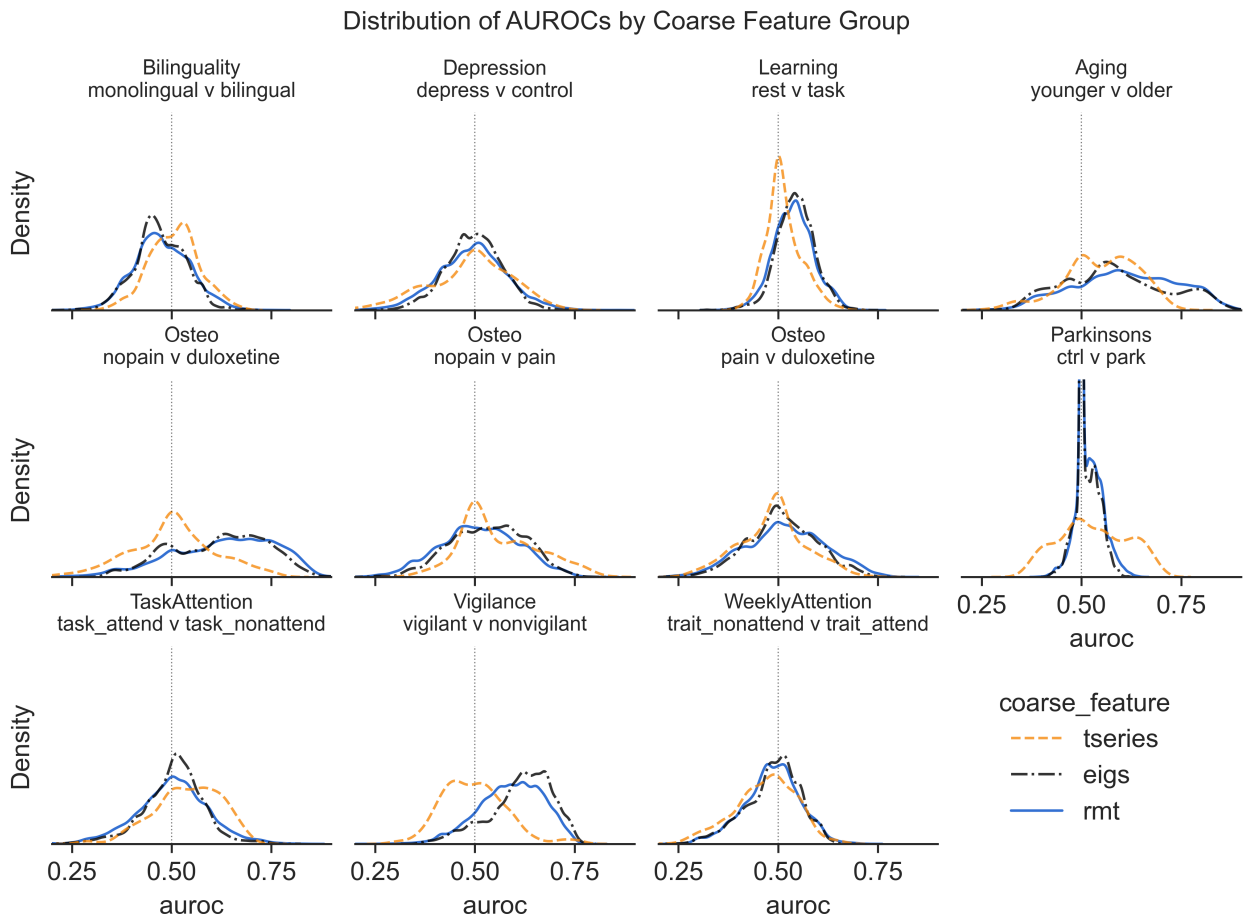

Fig 1 AUROC distributions across gross feature groupings and comparison tasks.

if, when restricting the analytic factor to a particular instance, there is no evidence of solvability. E.g. if a particular trimming procedure were to render all classification tasks unsolvable, it would be better to note this and exclude the associated mAUROCs from further visualizations, rather than to have the mAUROC distributions diluted by the bad procedure.

We take as lack of evidence of solvability an mAUROC distribution that is either (1) roughly symmetric and has mean and median close to 0.5, i.e. performance appears random, or (2) with median and mode less than 0.5. Based on the mAUROC distributions for either coarsely-grouped (Figure 1) or finely-grouped features (Figure S1), neither bilinguality nor depression were predictable by either baseline or eigenfeatures. Additionally, the `pain v duloxetine` comparison

in the *Osteo* dataset, and the `trait_attend v trait_nonattend` conditions ("WeeklyAttend" in Figures 1 and S1) in the *Attention* dataset were also not meaningfully predictable by any feature. As such, we exclude these classification tasks from further figures and discussion. We did not, however, find that any analytic factor resulted in unsolvability.

4.2.2 Predictable Comparisons

As visible in Figure 1, when coarsely summarizing features, eigenfeatures were more likely to be predictively useful than not, and except for in the `task_attend v task_nonattend` comparison, were also more predictively useful than the baseline features. Eigenfeatures most strongly and consistently demonstrated predictive utility in the *Aging* dataset `older v younger` classification task, the *Osteo* dataset `nopain v duloxetine` task, and in the *Attention* `vigilant v nonvigilant` comparison.

4.3 Largest mAUROCs

Considering the various analytic choices as tunable parameters, it makes sense to examine the *largest* portion of mAUROCs as an indication of the maximum predictive potential of the eigenfeatures. In this case, it is clear that eigenfeatures using RMT features almost always had the highest potential predictive utility (Figure 2). Figure S2 shows that this was primarily due to either the "rmt only" or "rmt + eigs" features (see Table 3). However, RMT eigenfeatures were also most likely to cause poor performance and overfitting (indicated by an mAUROC < 0.5; Figure S3).

Examining these RMT features more closely, it is clear these features' performance distributions differ mostly due to the unfolding procedure. That is, combined features that used the unfolded eigenvalues plus some other RMT eigenfeature tended to have visually-indistinguishable

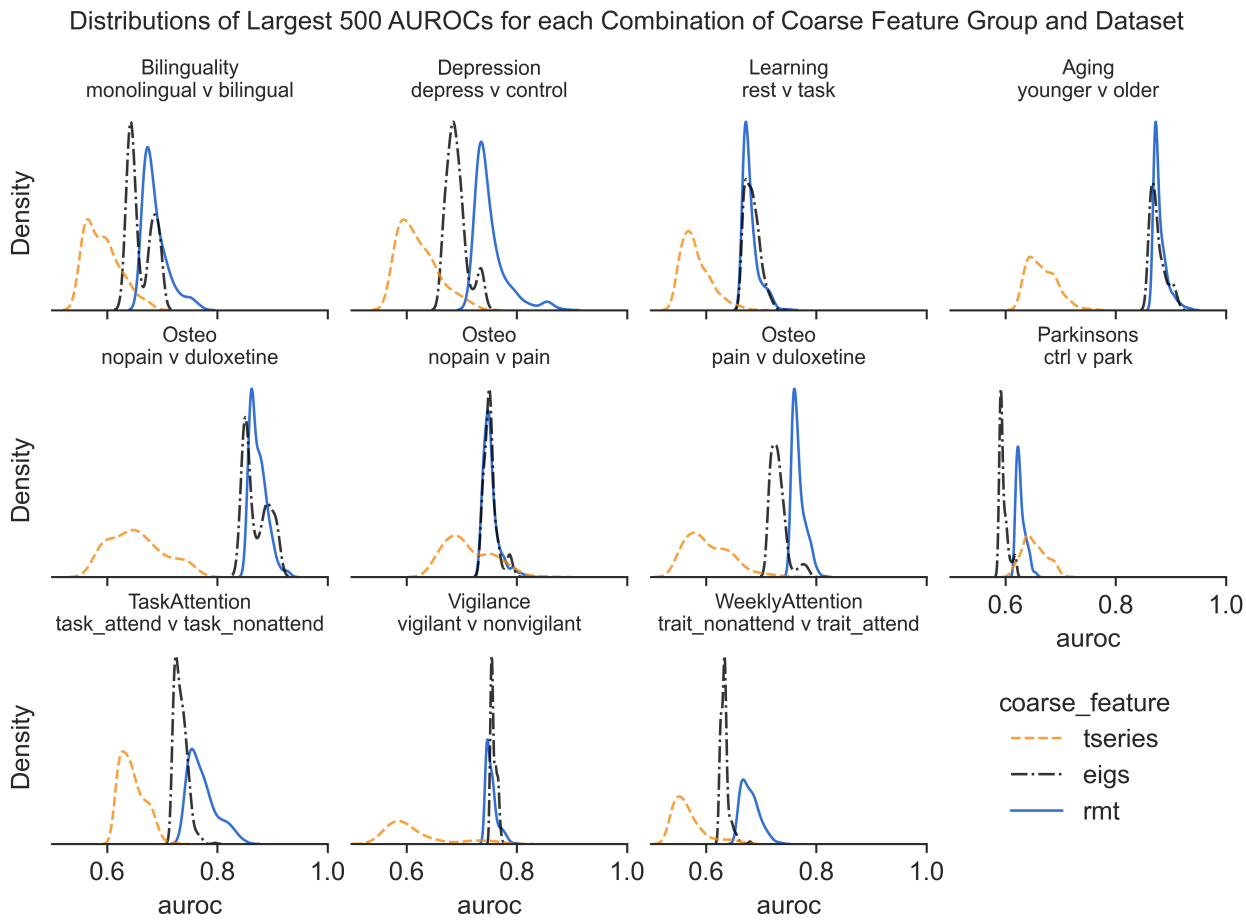

Fig 2 Distributions of largest 500 mAUROCs by coarse feature grouping.

mAUROC distributions to those using the unfolded eigenvalues alone (Figures S7-8). Instead, the mAUROC distributions of these features differed mostly in the tails (Figures S2-3).

4.4 Effect of Preprocessing

At a coarse level of feature grouping, slice time correction followed by motion correction tended to slightly increase the predictive utility of the eigenfeatures (Figure 3) relative to brain extraction only. Subsequent registration following these steps did not generally impact mAUROC distributions further, except in the `task_attend v task_nonattend` comparison, where registration reduced the predictive utility of the eigenfeatures (Figure 3, second-last column). When examin-

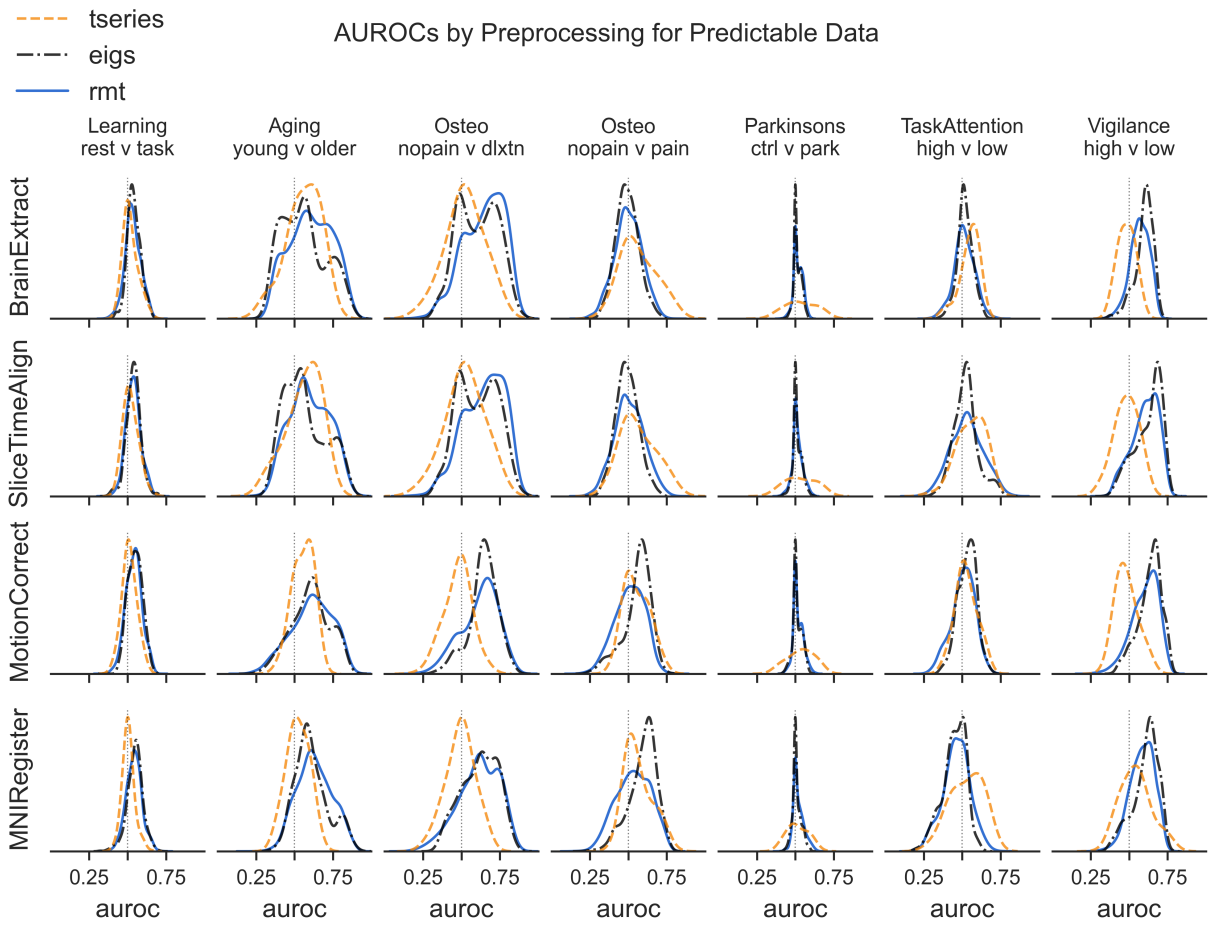

Fig 3 mAUROC distributions by preprocessing degree.

ing features more finely, it is clear that preprocessing most impacts the mAUROC distribution of the largest and central eigenvalues ("eigs middle" and "eigs max" in Figure S4).

4.5 Effect of Classifier

Within any feature grouping (coarse or fine), and within a classification task, mAUROC distributions were generally similar across classifiers (Figure S5, S6). Additionally, these figures also show that within each classification task, choice of classifier does not result in dramatic changes to the rough rank ordering of features. E.g. if features are ranked on predictive utility, using the median, modal, or bulk of the mAUROC values, this rank ordering appears to remain similar across

classifiers.

In the *Attention* `task_attend v task_nonattend` condition, eigenfeatures were modally predictive only with the random forest classifier, whereas in the *Aging* data, a SVC was least likely to have mAUROC < 0.5. Overall, however, differences in the mAUROC distributions due to classifier were small and inconsistent.

4.6 Normalization

There was no visible effect of feature normalization on mAUROC distributions.

4.7 Effect of Trimming

Trimming based on numerical precision (see Section 3.2.2 and Appendix B), did not result in meaningfully different mAUROC distributions in any case (Figure S7). However, trimming away the largest, or both largest and smallest eigenvalues, generally had a significant positive effect on the predictive quality of RMT features, most especially for RMT features involving the unfolded eigenvalues. When employing these trimming methods, these features were consistently more predictive than not (Figure S7).

4.8 Effect of Unfolding Degree

The choice of polynomial unfolding degree significantly impacted the mAUROC distributions for most classification tasks and most RMT features, and most significantly for the level variance features (Figure S8). Overall, Figure S8 weakly suggests that either smaller (degree = 3) or larger (degree = 9) unfolding degrees tend to yield the most predictively useful RMT features. However, when restricting to the most predictive RMT features (those including the unfolded eigenvalues),

it seems clear from Figure S8 that the largest unfolding degree of 9 produces the most favourable mAUROC distributions.

4.9 Effect of Slicing

The best slice size and location depended complexly on the classification task and feature, and few general summaries can be made of these interactions. However, features including the full spectrum (e.g. raw eigenvalues, smoothed eigenvalues, and their combinations with RMT eigenfeatures) were slightly more predictive when using the largest portions ("max-XX", rows in Figure S9), and usually least predictive when features primarily involved the smallest or central portions.

4.10 Choice of Summary Metric

We note briefly that most of the above findings regarding the impacts of analytic factors, and rank ordering of feature predictive utilities, are similar when using the adjusted accuracy (see Section 3.6), instead of the mAUROC (see Figures S10-18). However, if comparison task predictability is defined as requiring adjusted accuracies to be more positive than negative, then the *Learning* and *Parkinson's* datasets do not appear to be predictable with any feature (Figure S10).

5 Discussion

In this study, eigenfeatures inspired by RMT and derived from the eigenvalues of the full, whole-brain voxelwise fMRI correlation matrix were found to have predictive utility across a wide variety of phenomena and analytic choices. Compared to simple baseline reductions of the fMRI data, these eigenfeatures had more consistent predictive utility, and a higher maximum predictive potential (Figures 1-2). In addition to evidence from previous studies,[28,30,31] this suggests RMT may be a useful analytic and theoretical tool for understanding functional connectivity.

However, eigenfeature mAUROC values observed in this study were highly sensitive to the overall analytic procedure, and there was no single analytic choice (e.g. choice of trimming procedure, unfolding polynomial degree, number of preprocessing steps, or feature slicing) that ensured, for any combination of feature and classification task, that all other analytic choices resulted in mAUROC values greater than 0.5. In addition, the mean, median and modal mAUROCs were generally close to 0.5, and adjusted mean and median accuracies also tended to be close to zero. Thus we find limited evidence that functional-connectivity-based eigenfeatures have general, "out of the box" predictive utility, with general utility likely requiring either careful tuning, or different preprocessing decisions and analytic choices than those examined here.

Nevertheless, in all datasets, there were combinations of analytic choices that resulted in cross-validated mean prediction performances well beyond mere guessing (Figures S2, S11, Table 5). Whether or not these should be considered to have practical relevance depends on one's goals, however, we note that with small datasets of rs- or task-fMRI data, binary, subject-level classification using whole-brain features is generally challenging.

For example, deep learning methods improve upon guessing by 17% for autism[79] or 3-30% for ADHD,[80] 16% for severe depression,[81] and 23% for obsessive compulsive disorder.[82] Manual feature engineering with more separable conditions (e.g. schizophrenia) can result in classification accuracies well above 90%,[83] and with larger data, sophisticated custom feature extraction methods can achieve near perfect accuracies at classifying task vs. rest.[84] However, for functional connectivity data and with classical machine learning algorithms (such as SVC) we in general only expect large prediction accuracies (e.g. greater than 80%) when the group functional connectivities are already strongly separated (e.g. Cohen's $d > 1.0$).[85] In this study, the (robust) maximum improvements upon guessing are shown in Table 5, and vary from 3-26%.

Table 5 Top three robust maximum (95^{th} percentile) mAUROC and adjusted accuracy (acc+) values for each predictable classification task and fine feature grouping, sorted by mAUROC. A dash indicates that the fine feature grouping for that row was not in the top three, i.e., that the top three performing features differed depending on the performance metric.

Classification Task	Source Feature	mAUROC	acc+
Aging - younger v older	eigs smooth	0.834	0.226
	rmt + eigs	0.828	0.212
	eigs	0.823	0.211
Learning - rest v task	rmt + eigs	0.638	0.007
	eigs	0.636	0.007
	eigs max	0.631	–
	eigs middle	–	0.012
Osteo - nopain v duloxetine	rmt only	0.819	0.259
	eigs max	0.812	0.226
	rmt + eigs	0.806	–
	eigs smooth	–	0.212
Osteo - nopain v pain	tseries loc	0.762	0.104
	tseries scale	0.697	–
	eigs smooth	0.684	–
	eigs middle	–	0.069
	eigs	–	0.034
Parkinsons - ctrl v park	tseries scale	0.687	0.107
	tseries loc	0.657	0.063
	rmt only	0.590	0.027
TaskAttention - task_attend v task_nonattend	tseries loc	0.666	0.135
	rmt only	0.660	0.121
	tseries scale	0.656	0.124
Vigilance - vigilant v nonvigilant	eigs middle	0.733	0.184
	eigs smooth	0.726	–
	rmt + eigs	0.719	0.162
	eigs	–	0.162

It is somewhat surprising that the eigenfeatures examined here ever have net cross-validated predictive utility. The reduction of the functional connectivity matrix to the sorted $t-1$ eigenvalues uses all brain voxels (including grey matter voxels), and destroys a large amount of information (radically different matrices can have identical sorted eigenvalues). The subsequent small-degree polynomial fit used in the unfolding procedure further reduces variance in the raw data, and all eigenfeatures, due to the eigenvalue sorting, are monotonically increasing curves (or approximately monotonic). All such curves could likely be fit near-perfectly with 3-5 parameters, i.e. the inherent

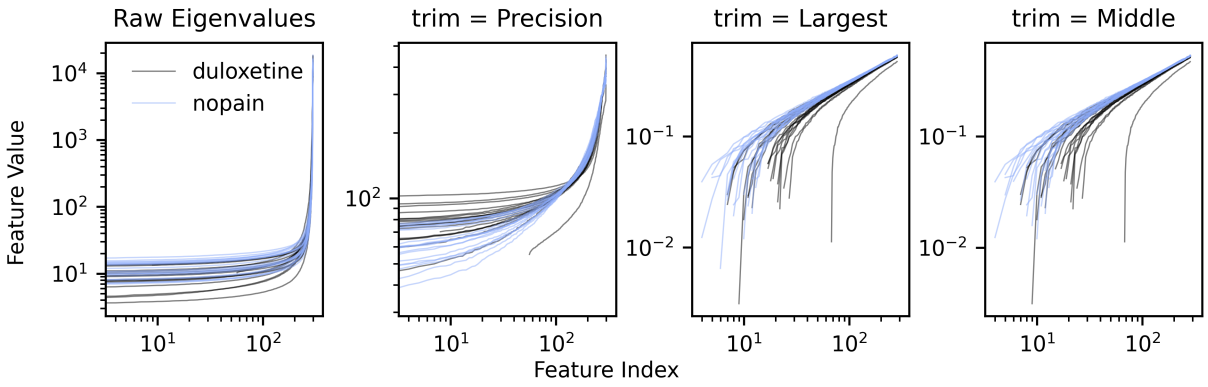

Fig 4 Raw eigenvalues (first subplot) and unfolded eigenvalues (polynomial degree 9; last 3 subplots) for *Osteo* dataset `duloxetine v nopain` classification task, brain extraction and slice time correction only. Median observed mAUROC = 0.643 (range = 0.346 - 0.925)

dimensionality of these features is quite small. Interpreting the eigenvalues as the magnitudes of the principal components of the standardized data, this suggests that a rough summary of the magnitudes of the principal components can often be surprisingly predictive in fMRI.

We speculate that the unfolded eigenvalues may have predictive utility in part because of their smoothing and re-scaling effect (see also Section 4.3). Figure 4, which depicts the raw eigenvalues and unfolded eigenvalues with different trimming procedures, shows how the raw eigenvalues have strongly exponential distributions, even with logarithmic axes. This is due to the magnitude of the largest eigenvalues, and the unfolded and trimmed feature distribution in Figure 4 are far less pathologically distributed.

5.1 Limitations

As is unfortunately typical of fMRI research,[86] the number of subjects in each dataset was quite small (Table 2). With such small numbers of subjects, randomization cannot be expected to effectively control group differences, and it is possible the predictive utility of the eigenfeatures was due to capitalization of such differences. For example, eigenfeatures were generally predictive in

the *Aging* dataset, and it is quite possible for randomization failures to introduce age imbalances across classes.

Likewise, one of the other more predictable classification tasks was the `duloxetine v nopain` task in the osteoarthritis dataset. An active drug could introduce any number of physiological confounds relevant to fMRI,[87] but we could not control for such effects do to the absence of physiological recording in most datasets.

In general, much larger fMRI datasets are needed to adequately control for and test *what* exactly the connectivity eigenvalues actually predict, and to test if these predictions generalize to larger, different populations.

6 Conclusion

Eigenvalue-based features inspired by RMT, and extracted from fMRI functional connectivity matrices were found to have predictive utility across a wide variety of datasets and classification tasks. However, the predictive utilities were modest, and highly dependent on preprocessing steps and other fitting and feature selection procedures. Future studies employing RMT should carefully investigate the sensitivity of any findings to such analytic decisions before drawing any conclusions. However, this study demonstrates that the inclusion of RMT statistics and other eigenfeatures extracted from fMRI images could aid in the predictive modeling of a wide variety of phenomena.

Disclosures

JL is founder of Time Will Tell Technologies, Inc.. The authors have no relevant financial interests in the manuscript and no other potential conflicts of interest to disclose.

Acknowledgments and Funding

This work was supported by the Natural Science and Engineering Research Council of Canada's Canada Research Chair grant (grant number 231266) to JL, Natural Science and Engineering Research Council of Canada Discovery Grant to JL, a Canada Foundation for Innovation and Nova Scotia Research and Innovation Trust infrastructure grant (R0176004) to JL, a St. Francis Xavier University research startup grant to JL (grant number R0168020), and a St. Francis Xavier University UCR grant to JL.

Data, Materials, and Code Availability

All raw fMRI data used in this study is publicly available on the OpenNeuro platform.[52,88]

All code used specifically to perform preprocessing and analyses for this paper is publicly available online in the paper GitHub repository.[89] This repository also includes all produced prediction data in a single table available in the file `all_produced_prediction_data.json.xz` available in the paper repository.

Code used to compute the spectral observables and perform unfolding is publicly available in the empyricalRMT library.[74]

Appendix A: Correlation Eigenvalues via Transposition

Let $\mathbf{X}$ be a real $n \times p$ matrix with $n > p$. Let

$$\mathbf{Z} = \text{norm}(\mathbf{X}) = \left(\mathbf{X}_1 - \bar{\mathbf{X}}_1 | \ldots | \mathbf{X}_p - \bar{\mathbf{X}}_p\right)$$

where $\mathbf{X}_i$ denotes column i of $\mathbf{X}$. Denote the (unordered) set of eigenvalues of $\mathbf{X}$ as $\text{eigs}(\mathbf{X})$, and let $r = (p-1)^{-1}$. Denote the covariance matrix of $\mathbf{X}$ as $\text{cov}(\mathbf{X})$. Then:

$$\begin{aligned}
\text{eigs}\left(\text{cov}(\mathbf{X})\right) &= \text{eigs}(r \cdot \mathbf{Z}\mathbf{Z}^\top) \\
&= r \cdot \text{eigs}(\mathbf{Z}\mathbf{Z}^\top) \\
&= r \cdot \text{eigs}((\mathbf{Z}\mathbf{Z}^\top)^\top) \\
&= r \cdot \text{eigs}(\mathbf{Z}^\top\mathbf{Z})
\end{aligned}$$

Denote the correlation matrix of $\mathbf{X}$ as $\text{corr}(\mathbf{X})$, and let

$$\mathbf{Y} = \text{standardize}(\mathbf{X}) = \left((\mathbf{X}_1 - \bar{\mathbf{X}}_1)/\sigma_1 | \ldots | (\mathbf{X}_p - \bar{\mathbf{X}}_p)/\sigma_p\right)$$

Then

$$\text{eigs}\left(\text{corr}(\mathbf{X})\right) = \text{eigs}\left(\text{cov}(\mathbf{Y})\right) = r \cdot \text{eigs}(\mathbf{Y}^\top\mathbf{Y})$$

Appendix B: Trimming Procedures

We implement three trimming procedures: precision-based, largest, and middle trimming. The source code is the definitive reference for the procedures, but we describe the motivations belows.

In precision-based trimming, we trim away any eigenvalues that are close enough to zero to be considered a result of numerical error due to floating point representation. There are two thresholds we consider, including the one used by NumPy[90] for the determination of matrix rank, and those recommended by LAPACK[91] in their user guide[92] on the error bounds for symmetric eigenproblems, and related additional details. We trim each matrix eigenvalues to whichever threshold is largest for the matrix in question.

For largest trimming, we must determine a threshold at which to separate "large" from "small" eigenvalues. The eigenvalues for our data tended to grow exponentially, so we instead looked at thresholding on the logarithms. We then used k-means with $k = 2$ on the precision-trimmed eigenvalues, and took the largest cluster (which also always had the smaller mean) cluster as the "largest" eigenvalues to trim away. "Middle" trimming simply reflects the threshold found by the largest trim method, e.g. if the largest trim method removes the last n precision-trimmed eigenvalues, then we also trim the first n smallest eigenvalues remaining after precision-trimming.

We chose one-dimensional k-means partly due to efficiency and simplicity, and because of the general relation to classical thresholding methods like the Otsu method.[93]

References

1 G. W. Anderson, A. Guionnet, and O. Zeitouni, *An Introduction to Random Matrices*, no. 118 in Cambridge Studies in Advanced Mathematics, Cambridge University Press, New York (2010).

2 G. Akemann, J. Baik, and P. Di Francesco, Eds., *The Oxford Handbook of Random Matrix Theory*, Oxford University Press, Oxford ; New York (2011).

3 M. L. Mehta, *Random Matrices*, no. v. 142 in Pure and Applied Mathematics, Academic Press, Amsterdam ; San Diego, CA, 3rd ed ed. (2004).

4 T. Guhr, A. Müller–Groeling, and H. A. Weidenmüller, "Random-matrix theories in quantum physics: Common concepts," *Physics Reports* **299**, 189–425 (1998).

5 T. A. Brody, J. Flores, J. B. French, *et al.*, "Random-matrix physics: Spectrum and strength fluctuations," *Reviews of Modern Physics* **53**, 385–479 (1981).

6 C. W. J. Beenakker, "Random-matrix theory of quantum transport," *Reviews of Modern Physics* **69**, 731–808 (1997).

7 O. Bohigas, R. U. Haq, and A. Pandey, "Higher-Order Correlations in Spectra of Complex Systems," *Physical Review Letters* **54**, 1645–1648 (1985).

8 D. Wintgen and H. Marxer, "Level statistics of a quantized cantori system," *Physical Review Letters* **60**, 971–974 (1988).

9 A. Pandey and S. Ghosh, "Skew-Orthogonal Polynomials and Universality of Energy-Level Correlations," *Physical Review Letters* **87**, 024102 (2001).

10 M. S. Santhanam and P. K. Patra, "Statistics of atmospheric correlations," *Physical Review E* **64**, 016102 (2001).

11 V. Plerou, P. Gopikrishnan, B. Rosenow, *et al.*, "Random matrix approach to cross correlations in financial data," *Physical Review E* **65**, 066126 (2002).

12 S. Jalan, C. Sarkar, A. Madhusudanan, *et al.*, "Uncovering Randomness and Success in Society," *PLoS ONE* **9**, e88249 (2014).

13 J. N. Bandyopadhyay and S. Jalan, "Universality in complex networks: Random matrix analysis," *Physical Review E* **76**, 026109 (2007).

14 S. Sastry, N. Deo, and S. Franz, "Spectral statistics of instantaneous normal modes in liquids and random matrices," *Physical Review E* **64**, 016305 (2001).

15 G. S. Matharoo, M. S. G. Razul, and P. H. Poole, "Spectral statistics of the quenched normal modes of a network-forming molecular liquid," *The Journal of Chemical Physics* **130**, 124512 (2009).

16 S. K. Sarkar, G. S. Matharoo, and A. Pandey, "Universality in the Vibrational Spectra of Single-Component Amorphous Clusters," *Physical Review Letters* **92**, 215503 (2004).

17 G. S. Matharoo, S. K. Sarkar, and A. Pandey, "Vibrational spectra of amorphous clusters: Universal aspects," *Physical Review B* **72**, 075401 (2005).

18 G. S. Matharoo, "Universality in the Vibrational Spectra of Amorphous Systems," *arXiv:0812.4613 [cond-mat]* (2008).

19 P. Bhadola and N. Deo, "Targeting functional motifs of a protein family," *Physical Review E* **94**, 042409 (2016).

20 I. Osorio and Y.-C. Lai, "A phase-synchronization and random-matrix based approach to multichannel time-series analysis with application to epilepsy," *Chaos: An Interdisciplinary Journal of Nonlinear Science* **21**, 033108 (2011).

21 A. Agrawal, C. Sarkar, S. K. Dwivedi, *et al.*, "Quantifying randomness in protein–protein interaction networks of different species: A random matrix approach," *Physica A: Statistical Mechanics and its Applications* **404**, 359–367 (2014).

22 A. Rai, A. V. Menon, and S. Jalan, "Randomness and preserved patterns in cancer network," *Scientific Reports* **4**, 6368 (2015).

23 S. B. Franklin, D. J. Gibson, P. A. Robertson, *et al.*, "Parallel Analysis: A method for determining significant principal components," *Journal of Vegetation Science* **6**(1), 99–106 (1995).

24 J. Veraart, D. S. Novikov, D. Christiaens, *et al.*, "Denoising of diffusion MRI using random matrix theory," *NeuroImage* **142**, 394–406 (2016).

25 M. Ulfarsson and V. Solo, "Dimension Estimation in Noisy PCA With SURE and Random Matrix Theory," *IEEE Transactions on Signal Processing* **56**, 5804–5816 (2008).

26 C. H. Martin and M. W. Mahoney, "Implicit Self-Regularization in Deep Neural Networks: Evidence from Random Matrix Theory and Implications for Learning," *Journal of Machine Learning Research* **22**(165), 1–73 (2021).

27 C. H. Martin, T. Peng, and M. W. Mahoney, "Predicting trends in the quality of state-of-the-art neural networks without access to training or testing data," *Nature Communications* **12**, 4122 (2021).

28 P. Šeba, "Random Matrix Analysis of Human EEG Data," *Physical Review Letters* **91**, 198104 (2003).

29 R. Wang, Z.-Z. Zhang, J. Ma, *et al.*, "Spectral properties of the temporal evolution of brain network structure," *Chaos: An Interdisciplinary Journal of Nonlinear Science* **25**, 123112 (2015).

30 R. Wang, L. Wang, Y. Yang, *et al.*, "Random matrix theory for analyzing the brain functional network in attention deficit hyperactivity disorder," *Physical Review E* **94**, 052411 (2016).

31 G. S. Matharoo and J. A. Hashmi, "Spontaneous back-pain alters randomness in functional connections in large scale brain networks: A random matrix perspective," *Physica A: Statistical Mechanics and its Applications* **541**, 123321 (2020).

32 A. M. Bastos and J.-M. Schoffelen, "A Tutorial Review of Functional Connectivity Analysis Methods and Their Interpretational Pitfalls," *Frontiers in Systems Neuroscience* **9** (2016).

33 M. P. van den Heuvel and H. E. Hulshoff Pol, "Exploring the brain network: A review on resting-state fMRI functional connectivity," *European Neuropsychopharmacology* **20**, 519–534 (2010).

34 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).

35 D. J. Lurie, D. Kessler, D. S. Bassett, *et al.*, "Questions and controversies in the study of time-varying functional connectivity in resting fMRI," *Network Neuroscience* **4**, 30–69 (2020).

36 M. Voultsidou, S. Dodel, and J. M. Herrmann, "Feature evaluation in fMRI data using random matrix theory," *Computing and Visualization in Science* **10**, 99–105 (2007).

37 A. A. Vergani, S. Martinelli, and E. Binaghi, "Resting state fMRI analysis using unsupervised learning algorithms," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* **0**, 1–14 (2019).

38 L. Gu, Z. Yu, T. Ma, *et al.*, "Random matrix theory for analysing the brain functional network in lower limb motor imagery," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 506–509 (2020).

39 S. E. Joel, B. S. Caffo, P. C. M. van Zijl, *et al.*, "On the relationship between seed-based and

ICA-based measures of functional connectivity," *Magnetic Resonance in Medicine* **66**(3), 644–657 (2011).

40 D. V. Smith, A. V. Utevsky, A. R. Bland, *et al.*, "Characterizing individual differences in functional connectivity using dual-regression and seed-based approaches," *NeuroImage* **95**, 1–12 (2014).

41 R. Wang, Z.-Z. Zhang, J. Ma, *et al.*, "Spectral properties of the temporal evolution of brain network structure," *Chaos: An Interdisciplinary Journal of Nonlinear Science* **25**, 123112 (2015).

42 A. Lo, H. Chernoff, T. Zheng, *et al.*, "Why significant variables aren't automatically good predictors," *Proceedings of the National Academy of Sciences* **112**, 13892–13897 (2015).

43 V. Amrhein, F. Korner-Nievergelt, and T. Roth, "The earth is flat ($p > 0.05$): Significance thresholds and the crisis of unreplicable research," *PeerJ* **5**, e3544 (2017).

44 T. Yarkoni and J. Westfall, "Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning," *Perspectives on Psychological Science* **12**, 1100–1122 (2017).

45 A. A. Abul-Magd and A. Y. Abul-Magd, "Unfolding of the spectrum for chaotic and mixed systems," *Physica A: Statistical Mechanics and its Applications* **396**, 185–194 (2014).

46 S. M. Abuelenin, "On the spectral unfolding of chaotic and mixed systems," *Physica A: Statistical Mechanics and its Applications* **492**, 564–570 (2018).

47 R. Fossion, G. T. Vargas, and J. C. L. Vieyra, "Random-matrix spectra as a time series," *Physical Review E* **88**, 060902 (2013).

48 S. M. Abuelenin and A. Y. Abul-Magd, "Effect of Unfolding on the Spectral Statistics of Adjacency Matrices of Complex Networks," *Procedia Computer Science* **12**, 69–74 (2012).

685 49 I. O. Morales, E. Landa, P. Stránský, *et al.*, "Improved unfolding by detrending of statistical
686 fluctuations in quantum spectra," *Physical Review E* **84**, 016203 (2011).
687 50 D. B. Parker and Q. R. Razlighi, "The Benefit of Slice Timing Correction in Common fMRI
688 Preprocessing Pipelines," *Frontiers in Neuroscience* **13** (2019).
689 51 J. P. Simmons, L. D. Nelson, and U. Simonsohn, "False-Positive Psychology: Undisclosed
690 Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant," *Psy-*
691 *chological Science* **22**, 1359–1366 (2011).
692 52 C. J. Markiewicz, K. J. Gorgolewski, F. Feingold, *et al.*, "The OpenNeuro resource for sharing
693 of neuroscience data," *eLife* **10**, e71774 (2021).
694 53 J. Fan, B. Mccandliss, J. Fossella, *et al.*, "The activation of attentional networks," *NeuroImage*
695 **26**, 471–479 (2005).
696 54 C. N. Wahlheim, A. P. Christensen, Z. M. Reagh, *et al.*, "Intrinsic Functional Connectiv-
697 ity in the Default Mode Network Predicts Mnemonic Discrimination: A Connectome-based
698 Modeling Approach," preprint, Neuroscience (2021).
699 55 C. E. Gold, A. L. Howell, J. Burdis, *et al.*, "Exploring the Resting State Neural Activity of
700 Monolinguals and Late and Early Bilinguals," (2019).
701 56 C. E. Gold, *Exploring the Resting State Neural Activity of Monolinguals and Late and Early*
702 *Bilinguals*. PhD thesis, Brigham Young University (2018).
703 57 B. DD, M. ME, S. AA, *et al.*, "Resting state with closed eyes for patients with depression and
704 healthy participants," (2021).
705 58 D. D. Bezmaternykh, M. Y. Melnikov, A. A. Savelov, *et al.*, "Brain Networks Connectivity in

Mild to Moderate Depression: Resting State fMRI Study with Implications to Nonpharmacological Treatment," *Neural Plasticity* **2021**, 1–15 (2021).

59 A. C. Schapiro, E. A. McDevitt, T. T. Rogers, *et al.*, "Human hippocampal replay during rest prioritizes weakly learned information and predicts memory performance," *Nature Communications* **9** (2018).

60 A. Schapiro, E. McDevitt, T. Rogers, *et al.*, "Human hippocampal replay during rest prioritizes weakly learned information and predicts memory performance," (2020).

61 P. Tétreault, A. Mansour, E. Vachon-Presseau, *et al.*, "Brain Connectivity Predicts Placebo Response across Chronic Pain Clinical Trials," *PLOS Biology* **14**, e1002570 (2016).

62 T. M. Madhyastha, M. K. Askren, P. Boord, *et al.*, "Dynamic Connectivity at Rest Predicts Attention Task Performance," *Brain Connectivity* **5**, 45–59 (2015).

63 K. J. Gorgolewski, N. Mendes, D. Wilfling, *et al.*, "A high resolution 7-Tesla resting-state fMRI test-retest dataset with cognitive and physiological measures," *Scientific Data* **2**, 140054 (2015).

64 K. J. Gorgolewski, D. Lurie, S. Urchs, *et al.*, "A Correspondence between Individual Differences in the Brain's Intrinsic Functional Architecture and the Content and Form of Self-Generated Thoughts," *PLoS ONE* **9**, e97176 (2014).

65 D. Watson and L. A. Clark, "The PANAS-X: Manual for the Positive and Negative Affect Schedule - Expanded Form," (1994).

66 L. Shalev, A. Ben-Simon, C. Mevorach, *et al.*, "Conjunctive Continuous Performance Task (CCPT)—A pure measure of sustained attention," *Neuropsychologia* **49**, 2584–2591 (2011).

67 V. Fonov, A. C. Evans, K. Botteron, *et al.*, "Unbiased average age-appropriate atlases for pediatric studies," *NeuroImage* **54**, 313–327 (2011).

68 V. Fonov, A. Evans, R. McKinstry, *et al.*, "Unbiased nonlinear average age-appropriate brain templates from birth to adulthood," *NeuroImage* **47**, S102 (2009).

69 R. Ciric, W. H. Thompson, R. Lorenz, *et al.*, "TemplateFlow: FAIR-sharing of multi-scale, multi-species brain models," preprint, Neuroscience (2021).

70 S. M. Smith, "Fast robust automated brain extraction," *Human Brain Mapping* **17**, 143–155 (2002).

71 B. B. Avants, N. J. Tustison, G. Song, *et al.*, "A reproducible evaluation of ANTs similarity metric performance in brain image registration," *NeuroImage* **54**, 2033–2044 (2011).

72 M. Jenkinson, P. Bannister, M. Brady, *et al.*, "Improved optimization for the robust and accurate linear registration and motion correction of brain images," *NeuroImage* **17**, 825–841 (2002).

73 M. Jenkinson, C. F. Beckmann, T. E. Behrens, *et al.*, "FSL," *NeuroImage* **62**, 782–790 (2012).

74 DM-Berger, "Stfxecutables/empyricalRMT: V1.1.1." Zenodo (2022).

75 F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research* **12**(85), 2825–2830 (2011).

76 S. Steegen, F. Tuerlinckx, A. Gelman, *et al.*, "Increasing Transparency Through a Multiverse Analysis," *Perspectives on Psychological Science* **11**, 702–712 (2016).

77 J. N. Mandrekar, "Receiver Operating Characteristic Curve in Diagnostic Test Assessment," *Journal of Thoracic Oncology* **5**, 1315–1316 (2010).

78 J. Brabec, T. Komárek, V. Franc, *et al.*, "On Model Evaluation under Non-constant Class Imbalance," (2020).

79 M. Bengs, N. Gessert, and A. Schlaefer, "4D Spatio-Temporal Deep Learning with 4D fMRI Data for Autism Spectrum Disorder Classification," *arXiv:2004.10165 [cs, eess]* (2020).

80 A. Riaz, M. Asad, E. Alonso, *et al.*, "DeepFMRI: End-to-end deep learning for functional connectivity and classification of ADHD using fMRI," *Journal of Neuroscience Methods* **335**, 108506 (2020).

81 R. Ramasubbu, M. R. G. Brown, F. Cortese, *et al.*, "Accuracy of automated classification of major depressive disorder as a function of symptom severity," *NeuroImage: Clinical* **12**, 320–331 (2016).

82 Y. Takagi, Y. Sakai, G. Lisi, *et al.*, "A Neural Marker of Obsessive-Compulsive Disorder from Whole-Brain Functional Connectivity," *Scientific Reports* **7**, 7538 (2017).

83 W. Du, V. D. Calhoun, H. Li, *et al.*, "High Classification Accuracy for Schizophrenia with Rest and Task fMRI Data," *Frontiers in Human Neuroscience* **6** (2012).

84 S. Zhang, X. Li, J. Lv, *et al.*, "Characterizing and differentiating task-based and resting state fMRI signals via two-stage sparse representations," *Brain Imaging and Behavior* **10**, 21–32 (2016).

85 C. Dansereau, Y. Benhajali, C. Risterucci, *et al.*, "Statistical power and prediction accuracy in multisite resting-state fMRI connectivity," *NeuroImage* **149**, 220–232 (2017).

86 B. O. Turner, E. J. Paul, M. B. Miller, *et al.*, "Small sample sizes reduce the replicability of task-based fMRI studies," *Communications Biology* **1**, 1–10 (2018).

87 K. Murphy, R. M. Birn, and P. A. Bandettini, "Resting-state FMRI confounds and cleanup," *NeuroImage* **80**, 349–359 (2013).

88 R. A. Poldrack and K. J. Gorgolewski, "OpenfMRI: Open sharing of task fMRI data," *NeuroImage* **144**, 259–261 (2017).

89 DM-Berger, "DM-Berger/random-matrix-fmri: Paper release." Zenodo (2022).

90 C. R. Harris, K. J. Millman, S. J. van der Walt, *et al.*, "Array programming with NumPy," *Nature* **585**, 357–362 (2020).

91 E. Anderson, Z. Bai, C. Bischof, *et al.*, *LAPACK Users' Guide*, Society for Industrial and Applied Mathematics, Philadelphia, PA, third ed. (1999).

92 E. Anderson, Z. Bai, C. Bischof, *et al.*, *LAPACK Users' Guide*, Society for Industrial and Applied Mathematics, third ed. (1999).

93 D. Liu and J. Yu, "Otsu Method and K-means," in *2009 Ninth International Conference on Hybrid Intelligent Systems*, **1**, 344–349 (2009).

Biographies and photographs of the other authors are not available.

List of Figures

List of Tables